

```

Db          33 PHPOFG 39

RESULT 6
Q92PK7      PRELIMINARY; PRT; 225 AA.
ID 092PK7
AC Q92PK7
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein SMC00528.
GN ORFNames=SMC00528;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiales; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;

RN
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolestad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godtke T., Goffeau A., Kahn D., Kles E., Lelaure V., Masny D.,
RA Pohl T., Pottecielie D., Puehler A., Purnelle B., Rampsberger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591788; CAC46316.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Set_estrs.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 225 AA; 25011 MW; C87CD753E13D92B CRC64;

Query Match          88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 7
Q8UED4      PRELIMINARY; PRT; 225 AA.
ID Q8UED4
AC Q8UED4; Q7CYE9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein Atcu1826 (AGR_C_3351P).
GN OrderedLocustNames=AGR_C_3351, Atcu1826;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiales; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;

RN
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Seubel J.C., Kaul R., Monks D.E., Kikajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. St.,
RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmach C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan M., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).

```

```

RN          [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Hailing C., Mullin L.,
RA Houtel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughny D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Giron V., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AB009137; AA14822.1; -.
DR EMBL; AB008102; AAK87595.1; -.
DR PIR; AH2800; AH2800.
DR PIR; B97580; B97580.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Set_estrs.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25084 MW; 1B7BDF5600841833 CRC64;

Query Match          88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 8
Q6FZW3      PRELIMINARY; PRT; 226 AA.
ID Q6FZW3
AC Q6FZW3;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=B005930;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;

RN
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
RA Alenmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,
RA Cabaack B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,
RA La Scola B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
DR EMBL; BX897700; CAP26085.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Set_estrs.
KW Complete proteome.
SQ SEQUENCE 226 AA; 25373 MW; D2AD7C07FBCB2625 CRC64;

Query Match          88.6%; Score 39; DB 2; Length 226;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 9
Q6G3B7      PRELIMINARY; PRT; 226 AA.
ID Q6G3B7
AC Q6G3B7;

```

DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=BH08660;
 OS Bartonella henselae (Rochalimaea henselae).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Bartonellaceae; Bartonella.
 NCBI_TaxID=38323;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 49882 / Houston 1;
 RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
 RA Almark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
 Candaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huet M.,
 La Scola B., Holmberg M., Andersson S.G.E.;
 RT "The louse-borne human pathogen Bartonella quintana is a genomic
 derivative of the zoonotic agent Bartonella henselae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
 DR EMBL: BX897699; CAF27664.1; -
 DR GO:GO:0003824; F:catalytic activity; IEA.
 DR InterPro: IPR000379; Set_sestr.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 226 AA; 25230 MW; 7216738444690A60 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 226;
 Best Local Similarity 85.7%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 :|||||
 DB 33 PHPQFG 39

RESULT 10
 Q98NR9 PRELIMINARY; PRT; 228 AA.
 AC Q98NR9;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE M110014 protein.
 GN OrderedLocusNames=m110014;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Rhizobiaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 Watanabe A., Ideesawa K., Ichikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP002994; BAB47692.1; -
 DR GO:GO:0003824; F:catalytic activity; IEA.
 DR InterPro: IPR000379; Set_sestr.
 KW Complete proteome.
 SQ SEQUENCE 228 AA; 25544 MW; B1A116C8EAE893BC CRC64;

Query Match 88.6%; Score 39; DB 2; Length 228;
 Best Local Similarity 85.7%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 :|||||
 DB 37 PHPQFG 43

RESULT 11
 Q96UF7 PRELIMINARY; PRT; 330 AA.
 AC Q96UF7;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE Subtilisin-like serine protease PRB (Fragment).
 GN Name=PRB;
 OS Metarhizium anisopliae var. acridum.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Ascomycetes; Clavicipitaceae;
 CC Microsporid Clavicipitaceae; Metarhizium.
 NCBI_TaxID=92637;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=FI-985;
 RA Bagga S., St. Leger R.T.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ416582; CAC95046.1; -
 DR HSSP; P06873; 2PRK.
 DR GO:GO:0008233; F:peptidase activity; IEA.
 DR GO:GO:0004289; F:subtilisin activity; IEA.
 DR GO:GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000209; Pept_S8_S53.
 DR InterPro: IPR009020; Prot_inh_propept.
 DR InterPro: IPR010259; Prot_inh_S8A.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Protease.
 FT NON TER 330 330
 SQ SEQUENCE 330 AA; 34673 MW; 09E294C2864E9130 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 330;
 Best Local Similarity 85.7%; Pred. No. 36;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 :|||||
 DB 149 SHPQFG 155

RESULT 12
 P511_SCHPO STANDARD; PRT; 379 AA.
 ID P511_SCHPO
 AC Q09912; Q9U92;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein p51 (Protein p51).
 GN Name=p51; Synonyms=ps1; ORFNames=SPCC830.07c;
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetes; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95290501; PubMed=7772606; DOI=10.1016/0167-4781(95)00063-M;
 RA Park S.-K., Chon S.-K., Yoo H.-S.;
 RT "A cDNA of Schizosaccharomyces pombe encoding a homologue of DnaJ-like
 protein.";
 RL Biochim. Biophys. Acta 1262:87-90(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=216848401; PubMed=11859360; DOI=10.1038/nature724;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowles E.,
 RA Rutherford K., Ruter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
 RA Wellens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabell C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wandut R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Nottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucera M., Rochet M., Galliardin C., Tallada V.A., Garçon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA "The genome sequence of *Schizosaccharomyces pombe*,"
 RT Nature 415:871-880(2002).
 CC -!- FUNCTION: Required for nuclear migration during mitosis. It is
 CC required for the normal initiation of translation.
 CC -!- SIMILARITY: Contains 1 J domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, L37753; AAA74732.1; -.
 DR EMBL, AL109850; CAB52880.1; -.
 DR PIR, S55900; S55900.
 DR PIR, T11633; T11633.
 DR HSSP, P25685; 1HDJ.
 DR GeneDB, SPombe; SPCC830.07c; -.
 DR InterPro; IPR002939; DnaJ_C.
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR008971; HSP40_DnaJ_pep.
 DR InterPro; IPR003095; Hsp_DnaJ.
 DR Pfam; PF00226; DnaJ_1.
 DR Pfam; PF01556; DnaJ_C_1.
 DR PRINTS; PR00625; DNAJPROTEIN.
 DR SMART; SM00371; DnaJ_1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS0076; DnaJ_2; 1.
 DR Cell cycle; Chapterone; DNA-binding.
 KW DOMAIN.
 FT DOMAIN 1 70 J-domain.
 FT CONFLICT 80 182 Gly-rich.
 FT CONFLICT 91 91 G -> C (in Ref. 1).
 SQ SEQUENCE 379 AA; 40260 MW; B8B949A1903F988A CRC64;

Query Match 88.6%; Score 39; DB 1; Length 379;
 Best Local Similarity 75.0%; Pred. No. 41;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPQFG 8
 Db 187 RSHPSFG 194

RESULT 13
 ID 096UP9 PRELIMINARY; PRT; 385 AA.
 AC 096UP9;
 DT 01-DEC-2001 (TRMBLrel. 19, Created)
 DT 01-DEC-2001 (TRMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)

DE Subtilisin-like protease PR1B.
 GN Name=PR1B;
 OS Metarrhizium anisopliae var. anisopliae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
 OC mitosporic Clavicipitaceae; Metarrhizium.
 OX NCBI_TaxID=92636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARSEF 820;
 RA Bagga S., St Leger R.J.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ416690; CAC95044.1; -.
 DR HSSP; P06873; 1IC6.
 DR GO; GO:0006233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000183; Decarboxylase.
 DR InterPro; IPR00209; Pept_S8_S53.
 DR InterPro; IPR009020; Pept_inh_propept.
 DR InterPro; IPR010259; Pept_inh_S8A.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Protease.
 SQ SEQUENCE 385 AA; 40094 MW; DB31BA0DC3A29FC3 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 385;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
 Db 149 THPQFG 155

RESULT 14
 ID 09P3Y1 PRELIMINARY; PRT; 385 AA.
 AC 09P3Y1;
 DT 01-OCT-2000 (TRMBLrel. 15, Created)
 DT 01-OCT-2000 (TRMBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
 DE Subtilisin-like protease PR1B.
 GN Name=PR1B;
 OS Metarrhizium anisopliae var. anisopliae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
 OC mitosporic Clavicipitaceae; Metarrhizium.
 OX NCBI_TaxID=92636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARSEF 2575;
 RA Bagga S., Screen S.E., St Leger R.J.;
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ289823; CAB95012.1; -.
 DR HSSP; P06873; 1IC6.
 DR GO; GO:0006233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000183; Decarboxylase.
 DR InterPro; IPR00209; Pept_S8_S53.
 DR InterPro; IPR009020; Pept_inh_propept.
 DR InterPro; IPR010259; Pept_inh_S8A.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Protease.
 SQ SEQUENCE 385 AA; 39959 MW; D9857DBEFC25D7C CRC64;

Query Match 88.6%; Score 39; DB 2; Length 385;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
 Db 149 THPQFG 155

RESULT 15

014410 PRELIMINARY; PRT; 386 AA.
 ID 014410;
 AC 014410;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Subtilisin-like protease PrtB (Fragment).
 GN Name=PrtB;
 OS Metarhizium anisopliae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
 OC Mitosporic Clavicipitaceae; Metarhizium.
 OX NCBI_TaxID=5530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARSEF 2575;
 RK MEDLINE=97473490; PubMed=9332344; DOI=10.1016/S0378-1119(97)00132-7;
 RA Joshi L., St Leger R.J., Roberts D.W.;
 RT "Isolation of a cDNA encoding a novel subtilisin-like protease (PrtB)
 RT from the entomopathogenic fungus, Metarhizium anisopliae using
 RT differential display-RT-PCR.";
 RL Gene 197:1-8(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARSEF 2575;
 RA Joshi L.S.T., Leger R.J., Roberts D.W.;
 RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL; U59484; AAC9831.1; -.
 DR HSP; P06873; 2PRK.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000183; Decarboxylase2.
 DR InterPro; IPR000209; Pept_S8_S53.
 DR InterPro; IPR009020; Prot_inh_propept.
 DR InterPro; IPR010259; Prot_inh_S8A.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PRO0723; SUBTILISIN.
 DR PROSITE; PS00879; ODR DC 2 2; UNKNOWN_1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Protease.
 FT NON_TER 386
 FT SEQUENCE 386 AA; 40031 MW; E805E0751C471568 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 386;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
 Db 150 THPQFG 156

Search completed: March 2, 2005, 12:44:18
 Job time : 30.7236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ7
Perfect score: 44
Sequence: 1 rxhpqfsg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq16Dec04:*

1: geneeqp1980s:*\n2: geneeqp1990s:*\n3: geneeqp2000s:*\n4: geneeqp2001s:*\n5: geneeqp2002s:*\n6: geneeqp2003as:*\n7: geneeqp2003bs:*\n8: geneeqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	8	6	ABP60360
2	39	88.6	9	2	AAW32485
3	39	88.6	9	2	AAW59211
4	39	88.6	9	2	AAW44010
5	39	88.6	9	2	AAW06914
6	39	88.6	9	4	AAW0795
7	39	88.6	9	5	ABG31053
8	39	88.6	9	5	AAW78476
9	39	88.6	9	6	AAO16086
10	39	88.6	9	6	ABP55547
11	39	88.6	9	6	AAE37229
12	39	88.6	9	6	AAE32860
13	39	88.6	9	6	AAE33270
14	39	88.6	9	6	ABG72479
15	39	88.6	9	7	ADB84587
16	39	88.6	9	7	ADJ34695
17	39	88.6	9	8	ADJ36219
18	39	88.6	9	8	ADJ88138
19	39	88.6	9	8	ADN29604
20	39	88.6	9	8	ADN62347
21	39	88.6	10	2	AAW66446
22	39	88.6	10	2	AAW50025
23	39	88.6	10	3	AAW57325
24	39	88.6	10	3	AAW67424
25	39	88.6	10	3	AAW19071

26	39	88.6	10	4	AAW61168	AAW61168	Paramagne
27	39	88.6	10	4	AAW99026	AAW99026	Streptavi
28	39	88.6	10	4	AAW02084	AAW02084	Streptavi
29	39	88.6	10	5	AAO14759	AAO14759	Human Int
30	39	88.6	10	7	ADD29929	ADD29929	Antibody-
31	39	88.6	11	2	AAW52691	AAW52691	PASK46-p1
32	39	88.6	11	2	AAW52692	AAW52692	PASK46-p1
33	39	88.6	12	2	AAW52696	AAW52696	PASK46-p1
34	39	88.6	12	2	AAW81821	AAW81821	Streptavi
35	39	88.6	18	5	AAW78473	AAW78473	Linker co
36	39	88.6	20	6	ADA37002	ADA37002	Artificia
37	39	88.6	23	5	AAO19984	AAO19984	Peptide o
38	39	88.6	42	5	AAO19988	AAO19988	Protein o
39	39	88.6	42	5	AAO19986	AAO19986	Protein o
40	39	88.6	46	8	ADP32686	ADP32686	Receptor
41	39	88.6	46	8	ADJ55658	ADJ55658	ED Peptid
42	39	88.6	47	2	AAW51892	AAW51892	VH1 Leade
43	39	88.6	47	5	AAW81776	AAW81776	Plasmid p
44	39	88.6	132	8	ADJ36263	ADJ36263	Self-coal
45	39	88.6	155	2	AAW32483	AAW32483	Kappa 11g

ALIGNMENTS

RESULT 1
ABP60360 standard, peptide, 8 AA.
ID ABP60360; XX
AC ABP60360; XX
XX 28-MAR-2003 (first entry)
DT XX
DE Streptavidin tag peptide SEQ ID NO 1.
XX Streptavidin; protein chip; microtitre plate; detection.
XX Synthetic.
XX DE10113776-A1.
XX 02-OCT-2002.
XX 21-MAR-2001; 2001DE-01013776.
XX 21-MAR-2001; 2001DE-01013776.
XX 21-MAR-2001; 2001DE-01013776.
XX (BIOA-) INST BIOANALYTIC GMBH GOETTINGEN.
XX Schmidt T;
XX WPI; 2003-031166/03.
XX New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.
XX Disclosure; Page 2; 18pp; German.
XX The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily controllable binding properties. (I) is particularly used for purifying FP from dilute solution in batch formats (which use simpler apparatus than column methods and result in lower loss of FP). The present sequence is that of a streptavidin tag peptide disclosed with the invention

```

XX SQ Sequence 8 AA;
      Query Match      88.6%; Score 39; DB 6; Length 8;
      Best Local Similarity 85.7%; Pred. NO. 1.8e+06;
      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 XHPQFG 8
DB 2 RHPQFG 8

RESULT 2
AAM32485
ID AAM32485 standard; peptide; 9 AA.
AC AAM32485;
XX
XX 27-MAR-1998 (first entry)
DT
XX
DE Strep-tag peptide.
XX
XX Catalytic antibody; growth factor; B-cell mitogenesis; LHL; L protein;
XX hen egg lysozyme; strep-tag; purification.
XX
XX OS Synthetic.
XX
XX WO9735887-A1.
XX
XX 02-OCT-1997.
XX
XX 26-MAR-1997; 97WO-AU000194.
XX
XX 26-MAR-1996; 96AU-00008951.
XX
XX 27-FEB-1997; 97AU-00005375.
XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Koentgen F., Suesse GM, Tarlinton DM, Treutlein HR;
XX
XX WPI; 1997-489572/45.
XX
XX New catalytic antibody precursors - comprising a B-cell surface molecule
XX binding portion which can induce B-cell mitogenesis.
XX
XX Example 3; Page 40; 109pp; English.
XX
XX This peptide comprises the strep-tag peptide that is recognised by
XX streptavidin. A form of novel growth factor LHL (see AAM32479) was
XX generated by PCR that contains a FLAG epitope (see AAM32484) at its N-
XX terminus and the strep-tag at its C-terminus. The construct is designated
XX LHL.seq (see AAM32481), where L is the immunoglobulin binding entity from
XX Peptostreptococcus magnus and H is residues 42-62 of hen egg lysozyme.
XX The strep-tag was used for purification of LHL.seq over a streptavidin
XX column. Thus, the LHL.seq was not purified on the basis of binding
XX immunoglobulin, thereby eliminating potential contamination by other
XX bacterial proteins which also bind immunoglobulins. LHL.seq has identical
XX activity to that of LHL, and can be used in novel methods for the
XX generation of catalytic antibodies
XX
XX Sequence 9 AA;
      Query Match      88.6%; Score 39; DB 2; Length 9;
      Best Local Similarity 85.7%; Pred. NO. 1.8e+06;
      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 XHPQFG 8
DB 3 RHPQFG 9

RESULT 3
AAM59211

```

```

ID AAM59211 standard; peptide; 9 AA.
XX
XX AAM59211;
AC
XX 27-AUG-1998 (first entry)
DT
XX
DE Streptavidin tagged peptide ligand #1.
XX
XX Streptavidin; ligand; binding affinity; mutant; isolation; purification;
XX recover; immobilise.
XX
XX OS Synthetic.
XX
XX EP835934-A2.
XX
XX 15-APR-1998.
XX
XX 09-OCT-1997; 97EP-00117504.
XX
XX 10-OCT-1996; 96DE-01041876.
XX
XX (BIOA-) INST BIOANALYTIK GMBH.
XX
XX Skerra A, Voss S;
XX
XX WPI; 1998-218868/20.
XX
XX Streptavidin mutants with higher binding affinity for peptide ligands -
XX have mutation in amino acid region 44-53, used to isolate, purify or
XX determine fusion proteins including these ligands.
XX
XX Claim 10; Page 11; 21pp; German.
XX
XX AAM59211 and AAM59212 are ligands used in a method to assay binding
XX affinity of streptavidin mutants. These mutants have a mutation within
XX the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX binding affinity than the wild-type for peptide ligands that include the
XX sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z
XX are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
XX mutants can be used to isolate, purify and determine proteins or to
XX determine/recover substances that contain streptavidin-binding groups.
XX Such compounds may also be used to immobilise fusions on microtitre
XX plates, microbeads or sensor chips
XX
XX Sequence 9 AA;
      Query Match      88.6%; Score 39; DB 2; Length 9;
      Best Local Similarity 85.7%; Pred. NO. 1.8e+06;
      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 XHPQFG 8
DB 3 RHPQFG 9

RESULT 4
AAM44010
ID AAM44010 standard; peptide; 9 AA.
AC AAM44010;
XX
XX 05-JUN-1998 (first entry)
DT
XX
DE Strep peptide epitope used in an epitope tagged prion protein construct.
XX
XX Prion; epitope; FLAG; Strep; poly-histidine; haemagglutinin; recombinant;
XX transgenic animal; scrapie; Creutzfeldt-Jakob disease; CJD;
XX bovine spongiform encephalopathy; BSE.
XX
XX OS Synthetic.
XX
XX WO9746572-A1.
XX

```

```

PD 11-DEC-1997;
XX
XX 29-MAY-1997; 97WO-US009289.
XX
XX 06-JUN-1996; 96US-00660626.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Prusiner SB, Telling GC, Cohen FE, Scott MR;
XX
XX WPI, 1998-042112/04.
XX
XX Nucleic acid construct encoding biologically active protein and epitope -
XX especially epitope-tagged prion protein.
XX
XX Claim 4, Page 48, 62pp; English.
XX
XX This sequence represents an artificial Strep peptide epitope. It is used
XX in a recombinant nucleic acid construct encoding an epitope-tagged prion
XX protein (PrP). The construct comprises a first nucleic acid sequence
XX encoding an amino acid sequence of a biologically active protein fragment
XX and a second nucleic acid sequence encoding a heterologous epitope
XX domain. The heterologous epitope domain is a peptide selected from a
XX peptide group of FLAG, Strep, poly-histidine, human c-myc peptide
XX recognised by monoclonal antibody 9B10 and haemagglutinin peptide
XX recognised by monoclonal antibody 12CA5. The protein is a natural,
XX synthetic or chimeric PrP molecule. The protein has two different three-
XX dimensional conformations and the epitope domain is spatially positioned
XX relative to the protein such that the epitope domain is more exposed in a
XX first conformation relative to a second conformation. The nucleic acid
XX construct may be used for the production of transgenic animals or cells
XX that are useful in a method for distinguishing between different
XX conformational shapes of a protein. These methods are particularly useful
XX in studying diseases caused by prion proteins, e.g. Creutzfeldt-Jakob
XX disease (CJD), scrapie and bovine spongiform encephalopathy (BSE)
XX
XX Sequence 9 AA;
SQ
Query Match 88.6%; Score 39; DB 2; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFG 8
DB 3 RHPQFG 9
RESULT 5
AAY06914
ID AAY06914 standard; peptide; 9 AA.
XX
XX AAY06914;
XX
XX 01-JUL-1999 (first entry)
XX
XX Strep-tag sequence.
XX
XX Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
XX catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
XX rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
XX human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
XX Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
XX autoimmune; inflammatory disease; gene therapy; strep-tag.
XX
XX Synthetic.
XX
XX WO915563-A1.
XX
XX 01-APR-1999.
XX
XX 18-SEP-1998; 98WO-AU000783.
XX
XX 19-SEP-1997; 97AU-00009306.

```

```

XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Koentgen F, Sness GM, Tarlinton DM, Treutlein HR;
XX
XX WPI, 1999-244394/20.
XX
XX Growth factor precursor cleaved by antigen-specific catalytic antibody.
XX
XX Example 3, Page 42; 101pp; English.
XX
XX The invention relates to a growth factor precursor that comprises B-cell
XX surface binding part, T cell surface binding part, antigen cleavable by a
XX catalytic antibody (CAB); and a peptide comprising heavy and light chains
XX of immunoglobulin. When the antigen is cleaved the B cell surface part
XX can interact with its target. The growth factor precursors are used to
XX select B cells that produce Ag-specific CAB, and to generate CAB from
XX such cells (by inducing mitogenesis, caused by the growth factor released
XX by specific cleavage). The Ag-specific CAB can be directed against, e.g.
XX tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
XX syndrome); viral docking receptors (treatment of human immune virus,
XX hepatitis and influenza infections); tumour-specific antigens; amyloid
XX plaque (treatment of Alzheimer's disease or myeloma) or IGE (treatment of
XX allergies such as asthma). CAB may also be used for drug detoxification,
XX to treat autoimmune or inflammatory diseases and to eliminate
XX environmental or industrial pollutants, such as plastics and petroleum.
XX Particularly the growth factor precursors are produced by delivering the
XX corresponding nucleic acid in a viral or other gene therapy vector. The
XX present sequence represents a strep-tag
XX
XX Sequence 9 AA;
SQ
Query Match 88.6%; Score 39; DB 2; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFG 8
DB 3 RHPQFG 9

```

```

RESULT 6
AAB30795
ID AAB30795 standard; peptide; 9 AA.
XX
XX AAB30795;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a Strep epitope.
XX
XX SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;
XX aggregation; fibril; phenotypic alteration; gene therapy;
XX disease resistance; plant pigmentation; prion disease.
XX
XX Synthetic.
XX
XX WO200075324-A2.
XX
XX 14-DEC-2000.
XX
XX 09-JUN-2000; 2000WO-US015876.
XX
XX 09-JUN-1999; 99US-0138833P.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
XX
XX WPI, 2001-061723/07.
XX
XX New nucleic acid encoding chimeric proteins with self-assembly
XX properties, useful e.g. for diagnosis and treatment of prion diseases,

```

PT also related aggregates, fibrils and polymers.
 XX
 PS Disclosure; Page 16; 188pp; English.
 CC
 CC The present sequence represents an epitope, which may be attached to
 CC chimeric polypeptides of the invention. The specification describes
 CC chimeric polypeptides which comprise at least one SCHAG (self-coalesces
 CC into higher-order aggregates) amino acid sequence fused in frame with a
 CC glutathione-S-transferase or a staphylococcal nuclear protein). The
 CC specification also describes chimeric polypeptides that comprises an
 CC amyloidogenic domain that causes aggregation into fibrils. The chimeric
 CC polypeptides are used to prepare polymers with multiple reactivities,
 CC e.g. derivatised with enzymes, or specific binding partners, and useful
 CC for performing multi-step chemical reactions. They can be used
 CC to create an inducible, or stable phenotypic alterations in a cell, e.g. for
 CC gene therapy, protein production, imparting disease resistance to plants,
 CC altering plant pigmentation and for diagnosis and treatment of prion
 CC diseases
 CC
 SO Sequence 9 AA;
 Query Match 88.6%; Score 39; DB 4; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 XHPQFG 8
 DB 3 RHPQFG 9
 RESULT 7
 ABG31053
 ID ABG31053 standard; peptide; 9 AA.
 XX
 AC ABG31053;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE Peptide production method associated strep-tag #1.
 XX
 KM Peptide production method; in vitro translation; in vitro transcription;
 KM strep-tag.
 XX
 OS Synthetic.
 XX
 PN WO200253582-A2.
 PD 11-JUL-2002.
 XX
 PF 06-DEC-2001; 2001WO-0P010682.
 PR 28-DEC-2000; 2000JP-00401417.
 PR 15-JAN-2001; 2001JP-00006910.
 PR 27-JUL-2001; 2001JP-00227094.
 PR 26-SEP-2001; 2001JP-00234795.
 PA (POST-) POST GENOME INST CO LTD.
 PI Inoue A, Shimizu Y, Ueda T;
 XX
 DR WPI; 2002-593608/64.
 XX
 PT Producing peptides, using reaction system for transcribing DNA into RNA
 PT and translating RNA, or for direct in vitro RNA translation, where
 PT protein components are labeled with a label pair adhering to each other.
 XX
 PS Disclosure; Page 44; 102pp; English.
 CC
 CC The invention describes a method of producing a peptide or its
 CC derivative, comprising using a reaction system for transcribing DNA into
 CC RNA and translating the RNA, or a reaction system for in vitro RNA
 CC translation, where a part or all of protein components constituting the

CC reaction system are labeled with one of a pair of substances adhering to
 CC each other and the other substance of the pair is used as an adsorbent
 CC for capturing labeled protein components after translation. The produced
 CC protein is efficiently isolated at a high purity from the reaction
 CC system, and at the same time, the problem of consumption of energy in the
 CC reaction system is resolved. This sequence represents a strep-tag which
 CC binds to streptavidin and can be used to purify peptides in the peptide
 CC production method of the invention
 CC
 SO Sequence 9 AA;
 Query Match 88.6%; Score 39; DB 5; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 XHPQFG 8
 DB 3 RHPQFG 9
 RESULT 8
 AAG78476
 ID AAG78476 standard; peptide; 9 AA.
 XX
 AC AAG78476;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Strep-tag nonapeptide linker sequence.
 XX
 KM Crystal lattice; crystallography; three dimensional structure;
 KM membrane protein; PMB908.
 XX
 OS Synthetic.
 XX
 PN WO200185962-A1.
 PD 15-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-GB002043.
 PR 05-MAY-2000; 2000SE-00001666.
 PR 02-JUN-2000; 2000US-0209331P.
 PR 28-JUN-2000; 2000SE-00002432.
 XX
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
 PA (IMAT/) IMATA S.
 PA (BYRN/) BYRNE B.
 PA (JORM/) JORMAKKA M.
 PA (ABRA/) ABRAMSON J.
 PA (SEJL/) SEJLITZ T.
 XX
 PI Iwata S, Byrne B, Jormakka M, Abramson J, Sejlitz T;
 XX
 DR WPI; 2002-089795/12.
 XX
 PT New recombinant vectors comprising promoter and nucleotide sequences,
 PT useful in methods of crystallization, particularly for the
 PT crystallization of proteins that are otherwise difficult to crystallize.
 XX
 PS Claim 19; Page 30; 70pp; English.
 CC
 CC This invention relates to recombinant vectors, comprising a promoter
 CC sequence and a nucleotide sequence encoding a first protein, which is a
 CC membrane protein, or multisubunit protein. The recombinant vector is
 CC useful in methods of crystallization. The vector is particularly useful
 CC for the crystallization of proteins that are otherwise difficult to
 CC crystallize. This sequence represents that are otherwise difficult to
 CC linker acts as a bridge between subunit IV and the foreign fusion protein
 CC of plasmid PMB908 (see AAH99982)
 CC
 SO Sequence 9 AA;

Query Match 88.6%; Score 39; DB 5; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 :|||||
 DB 3 RHPQFG 9

RESULT 9
 AAO16086
 ID AAO16086 standard; peptide; 9 AA.

AC AAO16086;

DT 27-FEB-2003 (first entry)

DE Neurological/CNS disease treatment method-related peptide #18.

XX Vaccine; gene therapy; neurological disease; CNS disorder;
 KW central nervous system disorder; olfactory system; Alzheimer's disease;
 KW Creutzfeld-Jakob disease; Huntington's chorea; Parkinson's disease;
 KW viral infection of the brain; brain tumour; lysosomal storage disease;
 XX multiple sclerosis.

OS Unidentified.

PN WO2002/4243-A2.

PD 26-SEP-2002.

PF 15-MAR-2002; 2002WO-US008042.

PR 15-MAR-2001; 2001US-00808037.

PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

PI (MCIN/) MCINNIS P.

PI Solomon B, Frenkel D;

DR WPI; 2003-040542/03.

PT Treating or diagnosing neurological diseases of the central nervous
 system, e.g. Alzheimer's disease, comprises displaying a polypeptide or
 PT diagnostic agent on viral display vehicle and introducing or detecting
 PT the display vehicle.

PS Disclosure; Page 214; 214pp; English.

XX The invention comprises a method for treating a neurological disease or a
 CC central nervous system (CNS) disorder. The method involves displaying a
 CC therapeutic molecule capable of treating the neurological disease or CNS
 CC disorder on a viral display vehicle. The viral display vehicle is then
 CC introduced into the olfactory system of a subject to treat the disease or
 CC disorder. The method of the invention is useful for preventing, treating
 CC and diagnosing neurological diseases or CNS disorders, such as:
 CC Alzheimer's disease; Creutzfeld-Jakob disease; Huntington's chorea; viral
 CC infections of the brain; brain tumours; lysosomal storage diseases;
 CC Parkinson's disease; and multiple sclerosis. The present amino acid
 CC sequence represents a peptide which was used in the invention

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 :|||||
 DB 3 RHPQFG 9

RESULT 10

ABP5547
 ID ABP5547 standard; peptide; 9 AA.

AC ABP5547;

DT 19-FEB-2003 (first entry)

DE Streptag peptide SEQ ID NO.77.

KW Hepatitis C virus; HCV; envelope protein; hepatotropic; virucide;
 KW antiinflammatory; gene therapy; vaccine.

XX Hepatitis C virus.
 OS Synthetic.

PN WO200285932-A2.

PD 31-OCT-2002.

PF 24-APR-2002; 2002WO-BE00062.

PR 24-APR-2001; 2001EP-00870088.

PR 17-JUL-2001; 2001US-0305604P.

PA (INNO-) INNOGENETICS NV.

PI Sablon E, Van Broekhoven A, Bosman A, Depla E, Deschamps G;

PI WPI; 2003-093095/08.

PT New recombinant nucleic acids for expressing Hepatitis C virus (HCV)
 PT envelope proteins in eukaryotic cells, comprising a sequence encoding a
 PT protein having an avian lysozyme leader peptide joined to the HCV
 PT envelope protein.

PS Claim 5; Page 303; 319pp; English.

XX The present invention describes a recombinant nucleic acid (I) comprising
 CC a nucleotide sequence encoding a protein having an avian lysozyme leader
 CC peptide, or its functional equivalent, joined to a Hepatitis C virus
 CC (HCV) envelope protein or its part. Also described: (1) a vector
 CC comprising the recombinant nucleic acid; (2) a host cell comprising the
 CC recombinant nucleic acid or the vector; and (3) a method for producing
 CC HCV envelope protein or its part in a host cell, comprising transforming
 CC the host cell with the recombinant nucleic acid or with the vector, where
 CC the host cell is capable of expressing the protein cited above. (I) has
 CC hepatotropic, virucide and antiinflammatory activities, and can be used
 CC in gene therapy and vaccines. The recombinant nucleic acid is useful for
 CC efficient expression of Hepatitis C virus envelope proteins in eukaryotic
 CC cells, such as yeast cells. The HCV envelope proteins may be used as a
 CC vaccine, or for incorporation into an immunoassay for the detection of
 CC anti-HCV antibodies, and/or genotyping of HCV, for prognosing or
 CC monitoring of HCV disease, or as a therapeutic agent. The method is used
 CC for producing HCV envelope protein or its part in a host cell. AB084197
 CC to AB084253 and ABP5528 to ABP5568 represent sequences used in the
 CC exemplification of the present invention

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 :|||||
 DB 3 RHPQFG 9

RESULT 11

AAE37229
 ID AAE37229 standard; peptide; 9 AA.

AC AAE37229;

```

XX 07-AUG-2003 (first entry)
XX Strep-tag peptide.
XX Gene expression; therapy; isolation.
XX Synthetic.
XX WO2003038049-A2.
XX 08-MAY-2003.
XX 29-OCT-2002; 2002WO-US034645.
XX 29-OCT-2001; 2001US-0340689P.
XX (RENO-) RENOVIS INC.
XX Heintz N, Serafini TA, Shyjan AW;
XX WPI; 2003-430512/40.
XX Isolating cell-type specific mRNA, useful in gene expression analysis or
XX quantification in a specific cell in a heterogeneous cell mixture, by
XX isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
XX type specific manner.
XX Disclosure; Page 8; 136pp; English.
XX The invention relates to a method for isolating mRNA from a population of
XX cells. The method involves selectively isolating ribosomes or proteins
XX that bind mRNA in a cell type specific manner and then isolating the mRNA
XX bound to the ribosomes or proteins that bind mRNA. The method is useful
XX for facilitating the analysis and quantification of gene expression in a
XX selected cell type present within a heterogeneous cell mixture. The
XX method may also be used in diagnostics or therapies for human diseases.
XX The present sequence is Strep-tag peptide. This sequence is used to
XX illustrate the method of the invention
XX
XX Sequence 9 AA;
SQ
Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 0;
QY 2 XHPQFGG 8
Db 3 RHPQFGG 9
RESULT 12
AAE32860
ID AAE32860 standard; peptide; 9 AA.
XX
XX AAE32860;
AC
XX
XX 24-MAR-2003 (first entry)
DT
XX
XX Strep-tag peptide.
DE
XX Hepatitis C virus; HCV; envelope protein; virucide; immunostimulant;
XX immune response; T-cell; therapy; infection; pharmaceutical; vaccine.
XX
XX Synthetic.
XX OS
XX WO200286101-A2.
XX PN
XX 31-OCT-2002.
XX PD
XX 24-APR-2002; 2002WO-BE000064.
XX PF
XX 24-APR-2001; 2001EP-00870088.
XX PR

```

```

PR 17-JUL-2001; 2001US-0305604P.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Depia E, Bosman A, Deschamps G, Sablon E, Suckow M, Samson I;
XX Verheyden G;
XX WPI; 2003-103409/09.
XX
XX New Hepatitis C virus (HCV) envelope protein with N-glycosylation
XX site(s), useful as a vaccine for inducing a HCV-specific immune response
XX or HCV-specific antibodies, particularly for preventing or treating HCV
XX infection.
XX
XX Claim 10; Page 336; 355pp; English.
XX
XX The invention relates to Hepatitis C virus (HCV) envelope protein with N-
XX glycosylation site(s). The HCV envelope protein, or the pharmaceutical
XX composition comprising the envelope protein, is useful as a medicament or
XX a vaccine, particularly for inducing a HCV-specific immune response,
XX inducing HCV-specific antibodies or inducing a T-cell function in a
XX mammal. The protein is particularly useful for preventing, treating or
XX diagnosing HCV infection. It is also useful for detecting the presence of
XX anti-HCV antibodies in a sample. The present sequence is a peptide used
XX in the invention
XX
XX Sequence 9 AA;
SQ
Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 0;
QY 2 XHPQFGG 8
Db 3 RHPQFGG 9
RESULT 13
AAE33270
ID AAE33270 standard; peptide; 9 AA.
XX
XX AAE33270;
AC
XX
XX 02-APR-2003 (first entry)
DT
XX
XX Strep epitope tag used to illustrate the method of the invention.
XX
XX Membrane-spanning signal-transducing protein; MSST protein; epitope.
XX
XX Synthetic.
XX OS
XX WO200286507-A1.
XX PN
XX 31-OCT-2002.
XX PD
XX 24-APR-2002; 2002WO-US013250.
XX PF
XX 24-APR-2001; 2001US-0286250P.
XX PR
XX 21-AUG-2001; 2001US-00935061.
XX
XX (STRD ) UNIV DELAND STANFORD JUNIOR.
XX PA
XX Kobilka BK, Ghanouni P, Lee TW;
XX PI
XX WPI; 2003-103418/09.
XX DR
XX
XX Identifying an agent that modulates activity of a membrane-spanning,
XX signal-transducing (MSST) protein, by detecting a conformational change
XX in a MSST protein upon interaction with a ligand.
XX
XX Disclosure; Page 84; 104pp; English.
XX
XX The present invention relates to methods and compositions for identifying
XX

```

CC Agents that modulate activity of a membrane-spanning, signal-transducing
 CC (MST) protein. The method involves detecting a conformational change in
 CC a MST protein upon interaction with a ligand. The method is useful for
 CC identifying agents that modulate (e.g. agonists or antagonists) activity
 CC of MST protein. The present sequence is a strep epitope tag used to
 CC illustrate the method of the invention

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
 Db 3 RHPQFG 9

RESULT 14

ABG72479 standard; peptide; 9 AA.

AC ABG72479;

DT 18-FEB-2003 (first entry)

DE Strep tag for use as a G protein epitope tag.

XX G protein; alpha sub-unit; G protein coupled receptor; GPCR;

KW G protein activation; G protein coupled receptor activation;

KW G protein coupled receptor identification; strep tag.

OS Synthetic.

PN US6448377-B1.

PD 10-SEP-2002.

PF 27-SEP-2000; 2000US-00672239.

PR 27-SEP-2000; 2000US-00672239.

PA (STRD) UNIV LEIAND STANFORD JUNIOR.

PI Koblika B, Lee TW;

DR WPI; 2003-110149/10.

PT Determining effects of candidate agent on activation of a G protein
 PT coupled receptor (GPCR) for evaluating new agonists and/or inverse
 PT agonists for GPCRs by contacting a candidate agent with a modified G
 PT protein alpha subunit and a GPCR.

PS Disclosure; Col 9; 38pp; English.

XX The invention describes a method of determining the effects of a
 CC candidate agent on activation of a G protein coupled receptor (GPCR). The
 CC method comprises contacting a candidate agent with a modified G protein
 CC alpha subunit and a GPCR, and detecting a level of G protein activation
 CC in response to the contacting, where the level of activation is
 CC indicative of the effects of the agent on the activity of GPCR. The
 CC method is useful for determining the effects of a candidate agent on
 CC activation of a G protein coupled receptor, evaluating new agonists,
 CC and/or inverse agonists for GPCRs, identifying ligands for GPCRs, and
 CC developing a strategy for identifying GPCRs involved in different
 CC biological processes, including diseases. The invention provides rapid
 CC and more sensitive bioassays for evaluating new agonists, agonists and/or
 CC inverse agonist for GPCRs. The method can be performed using membranes,
 CC which increases both the ease of performing the assay and its efficacy,
 CC and also allows high throughput screening of GPCR activity. Furthermore,
 CC this method directly measures GPCR activity, and thus is less labour-
 CC intensive than the conventional methods. This sequence represents the
 CC Strep tag, an example of an epitope tag that can be used to tether the

CC stimulatory G protein alpha sub-unit to a membrane

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
 Db 3 RHPQFG 9

RESULT 15

ADB84587 standard; peptide; 9 AA.

AC ADB84587;

DT 04-DEC-2003 (first entry)

DE Streptavidin conserved peptide #1.

XX cell-free transcription system; cell-free translation system;

KW protein synthesis; matrix; streptavidin.

OS Escherichia coli.

PN DE10137792-A1.

PD 27-FEB-2003.

PF 06-AUG-2001; 2001DE-01037792.

PR 06-AUG-2001; 2001DE-01037792.

PA (ERDM/) ERDMANN V.

PI Erdmann VA, Lamla T, Stiege W;

DR WPI; 2003-343999/33.

PT Expressing genes in cell-free system, useful for preparation of proteins,
 PT comprises that the protein formed is removed from solution by binding to
 PT a matrix.

PS Claim 13; Col 8; 8pp; German.

XX This invention describes a novel method of expressing genes in a cell-
 CC free transcription and translation system which comprises using a
 CC reaction solution containing all necessary components of the
 CC transcription/translation system, amino acids, nucleotides and
 CC metabolites that supply energy and that are needed for synthesis. The
 CC proteins formed are immobilised on a matrix. The method allows simple
 CC recovery of proteins without a separate isolation step and the amount of
 CC proteins produced can be determined before a reaction is complete.

CC Continuous removal of proteins prevents it interfering with the
 CC expression process, making possible synthesis of proteins that interact
 CC adversely with the process, so normally produced only in very low yields.
 CC The use of a matrix concentrates the proteins produced and is applicable
 CC to proteins of any size. This sequence represents a highly conserved
 CC peptide from E. coli streptavidin which is used to illustrate the method
 CC of the invention.

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 7; Length 9;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
 Db 3 RHPQFG 9

Wed Mar 2 14:27:15 2005

seq7.rag

Page 8

Search completed: March 2, 2005, 13:02:45
Job time : 36.0325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ7
Perfect score: 44
Sequence: 1 rxhpqfgg 8

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 1365339 seqs, 328044528 residues
Total number of hits satisfying chosen parameters: 1365339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	8	US-10-026-578B-1	Sequence 1, Appl
2	39	88.6	9	US-09-808-037-32	Sequence 32, Appl
3	39	88.6	9	US-09-983-067-2	Sequence 2, Appl
4	39	88.6	9	US-09-935-061-3	Sequence 3, Appl
5	39	88.6	9	US-10-128-590-77	Sequence 77, Appl
6	39	88.6	9	US-10-345-618-14	Sequence 14, Appl
7	39	88.6	9	US-10-128-587A-77	Sequence 77, Appl
8	39	88.6	9	US-10-384-788-32	Sequence 32, Appl
9	39	88.6	9	US-10-449-831A-166	Sequence 166, App
10	39	88.6	9	US-10-425-000-75	Sequence 75, Appl
11	39	88.6	9	US-10-424-999-24	Sequence 24, Appl
12	39	88.6	9	US-10-272-196-30	Sequence 30, Appl
13	39	88.6	9	US-10-612-410-18	Sequence 18, Appl

14	39	88.6	9	US-10-692-071-3	Sequence 3, Appl
15	39	88.6	9	US-10-397-438A-6	Sequence 6, Appl
16	39	88.6	9	US-10-753-309-5	Sequence 5, Appl
17	39	88.6	9	US-10-494-248-5	Sequence 18, Appl
18	39	88.6	10	US-09-801-968-18	Sequence 18, Appl
19	39	88.6	10	US-09-802-154-18	Sequence 13, Appl
20	39	88.6	10	US-10-060-765-13	Sequence 12, Appl
21	39	88.6	10	US-10-263-230A-12	Sequence 13, Appl
22	39	88.6	10	US-10-818-140-13	Sequence 13, Appl
23	39	88.6	10	US-10-771-173-13	Sequence 8, Appl
24	39	88.6	12	US-09-904-599A-8	Sequence 9, Appl
25	39	88.6	18	US-10-397-438A-9	Sequence 1, Appl
26	39	88.6	20	US-10-342-805-1	Sequence 2, Appl
27	39	88.6	23	US-10-344-607-2	Sequence 16, Appl
28	39	88.6	42	US-10-344-607-16	Sequence 20, Appl
29	39	88.6	46	US-10-344-607-20	Sequence 12, Appl
30	39	88.6	46	US-10-422-262-12	Sequence 8, Appl
31	39	88.6	46	US-10-448-609-8	Sequence 57, Appl
32	39	88.6	47	US-10-416-290-57	Sequence 212, App
33	39	88.6	132	US-10-449-831A-212	Sequence 11, Appl
34	39	88.6	155	US-10-345-618-11	Sequence 13, Appl
35	39	88.6	178	US-10-345-618-13	Sequence 13, Appl
36	39	88.6	194	US-10-416-708A-10	Sequence 64, Appl
37	39	88.6	194	US-10-416-708A-64	Sequence 8, Appl
38	39	88.6	198	US-10-345-618-8	Sequence 24, Appl
39	39	88.6	210	US-09-272-809-24	Sequence 198, App
40	39	88.6	212	US-10-449-831A-198	Sequence 192, App
41	39	88.6	234	US-10-449-831A-192	Sequence 204, App
42	39	88.6	329	US-10-449-831A-204	Sequence 4, Appl
43	39	88.6	334	US-10-017-736-4	Sequence 4, Appl
44	39	88.6	334	US-10-650-585-4	Sequence 6, Appl
45	39	88.6	342	US-10-345-618-6	

ALIGNMENTS

RESULT 1
US-10-026-578B-1
Sequence 1, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity To
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-1

Query Match 88.6%; Score 39; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFG 8
DB 2 RHPQFG 8

```
RESULT 2
US-09-808-037-32
; Sequence 32, Application US/09808037
; Patent No. US2002005231A1
; GENERAL INFORMATION:
; APPLICANT: SOLOMON, Beka
; APPLICANT: HANAN, Elia
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT AND/OR DIAGNOSIS OF
; TITLE OF INVENTION: NEUROLOGICAL DISEASES AND DISORDERS
; FILE REFERENCE: SOLOMON-2D
; CURRENT APPLICATION NUMBER: US/09/808,037
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 09/629,971
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/473,653
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: US 60/152,417
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 9
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-808-037-32

Query Match
Best Local Similarity 88.6%; Score 39; DB 9; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 3
US-09-983-067-2
; Sequence 2, Application US/09983067
; Patent No. US2002012310A1
; GENERAL INFORMATION:
; APPLICANT: INOUE, Akio
; APPLICANT: SHIMIZU, Yoshihiro
; APPLICANT: UEDA, Takuya
; TITLE OF INVENTION: Process for Producing Peptides By Using In Vitro
; FILE REFERENCE: 1752-0151P
; CURRENT APPLICATION NUMBER: US/09/983,067
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: JP 294795/2001
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: JP 227094/2001
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: JP 6910/2001
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 401417/2000
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strep-tag binding to streptavidin
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(9)
; OTHER INFORMATION: Schmidt & Skeris, 1993, "The random peptide library-assisted
; OTHER INFORMATION: engineering of a C-terminal affinity peptide, useful for the
; OTHER INFORMATION: detection and purification of a functional IgFv fragment",
; OTHER INFORMATION: Protein Eng. 6(1):109-122.
US-09-983-067-2
```

```
Query Match
Best Local Similarity 88.6%; Score 39; DB 9; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 4
US-09-935-061-3
; Sequence 3, Application US/09935061
; Publication No. US20030123649A1
; GENERAL INFORMATION:
; APPLICANT: Koblika, Brian M.
; APPLICANT: Ghanouni, Pejman
; APPLICANT: Lee, Tae Weon
; TITLE OF INVENTION: Conformational assays to detect binding
; TITLE OF INVENTION: to G protein-coupled receptors
; FILE REFERENCE: STAN213
; CURRENT APPLICATION NUMBER: US/09/935,061
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/286,250
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: epitope tag peptide
US-09-935-061-3

Query Match
Best Local Similarity 88.6%; Score 39; DB 10; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 5
US-10-128-590-77
; Sequence 77, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Immunogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 9
; TYPE: PRT
; ORGANISM: streptag
US-10-128-590-77

Query Match
Best Local Similarity 88.6%; Score 39; DB 14; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 6
US-10-345-618-14
; Sequence 14, Application US/10345618
```

```
/ Publication No. US2003014848A1
/ GENERAL INFORMATION:
/ APPLICANT: Koenigsen, Frank
/ APPLICANT: Sues, Gabriele M.
/ APPLICANT: Tarlinton, David M.
/ APPLICANT: Treutlein, Herbert R.
/ TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF PRODUCING SAME
/ FILE REFERENCE: 13474
/ CURRENT APPLICATION NUMBER: US/10/345,618
/ CURRENT FILING DATE: 2003-01-16
/ PRIOR APPLICATION NUMBER: US/09/509,031
/ PRIOR FILING DATE: 2000-06-09
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-10-345-618-14

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 7
US-10-128-587A-77
/ Sequence 77, Application US/10128587A
/ Publication No. US20030152940A1
/ GENERAL INFORMATION:
/ APPLICANT: Imogenetics N.V.
/ TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
/ FILE REFERENCE: 134 PCT
/ CURRENT APPLICATION NUMBER: US/10/128,587A
/ CURRENT FILING DATE: 2002-04-24
/ NUMBER OF SEQ ID NOS: 98
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 77
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: streptag
US-10-128-587A-77

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 8
US-10-384-788-32
/ Sequence 32, Application US/10384788
/ Publication No. US20040013647A1
/ GENERAL INFORMATION:
/ APPLICANT: SOLOMON, Bekk
/ APPLICANT: FRENKEL, Dan
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING A PLAQUE-FORMING DISEASE
/ FILE REFERENCE: SOLOMON=2D.2
/ CURRENT APPLICATION NUMBER: US/10/384,788
/ CURRENT FILING DATE: 2003-03-11
/ PRIOR APPLICATION NUMBER: 60/371,735
```

```
/ PRIOR FILING DATE: 2002-04-12
/ PRIOR APPLICATION NUMBER: 09/808,037
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: 09/830,954
/ PRIOR FILING DATE: 2001-06-22
/ PRIOR APPLICATION NUMBER: 10/162,889
/ PRIOR FILING DATE: 2002-06-06
/ PRIOR APPLICATION NUMBER: 09/473,653
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: 09/629,971
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: 60/152,417
/ PRIOR FILING DATE: 1999-09-03
/ PRIOR APPLICATION NUMBER: PCT/IL00/00518
/ PRIOR FILING DATE: 2000-08-31
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 32
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Synthetic construct
US-10-384-788-32

Query Match      88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 9
US-10-449-831A-166
/ Sequence 166, Application US/10449831A
/ Publication No. US20040029179A1
/ GENERAL INFORMATION:
/ APPLICANT: Koenigsen, Frank
/ TITLE OF INVENTION: Higher molecular weight entities and uses therefor
/ FILE REFERENCE: 2385978
/ CURRENT APPLICATION NUMBER: US/10/449,831A
/ CURRENT FILING DATE: 2003-05-30
/ PRIOR APPLICATION NUMBER: USSN 60/384878
/ PRIOR FILING DATE: 2002-05-31
/ NUMBER OF SEQ ID NOS: 237
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 166
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Strep tag
US-10-449-831A-166

Query Match      88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 10
US-10-425-000-75
/ Sequence 75, Application US/10425000
/ Publication No. US20040052777A1
/ GENERAL INFORMATION:
/ APPLICANT: Nesbitt, Mark
/ APPLICANT: Cameron, Beatrice
/ APPLICANT: Blanche, Francis
/ TITLE OF INVENTION: Kringles Polypeptides and Methods for Using Them to Inhibit
/ TITLE OF INVENTION: Angiogenesis
```

```
FILE REFERENCE: ST01027-B
CURRENT APPLICATION NUMBER: US/10/425,000
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: Patentin version 3.2
SEQ ID NO 75
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-425-000-75

Query Match
Best Local Similarity 88.6%; Score 39; DB 15; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 11
US-10-424-999-24
Sequence 24, Application US/10/424,999
GENERAL INFORMATION:
APPLICANT: Nesbitt, Mark
APPLICANT: Cameron, Beatrice
TITLE OF INVENTION: Using Them to Inhibit Angiogenesis
FILE REFERENCE: ST01027-A
CURRENT APPLICATION NUMBER: US/10/424,999
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 70
SOFTWARE: Patentin version 3.2
SEQ ID NO 24
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-424-999-24

Query Match
Best Local Similarity 88.6%; Score 39; DB 15; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 12
US-10-272-196-30
Sequence 30, Application US/10/272,196
GENERAL INFORMATION:
APPLICANT: Murray, Clare Margaret
APPLICANT: Hutchinson, Raymond
APPLICANT: Bantick, John Raymond
APPLICANT: Sullivan, Michael
APPLICANT: Donald, David Keith
APPLICANT: Jackson, Andrew Paul
APPLICANT: Jackson, Clive Geoffrey
APPLICANT: Cook, Ian David
TITLE OF INVENTION: INHIBITORS OF MONOCARBOXYLATE TRANSPORT
FILE REFERENCE: 06275-285001
```

```
CURRENT APPLICATION NUMBER: US/10/272,196
CURRENT FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: US 60/329,318
PRIOR FILING DATE: 2001-10-16
NUMBER OF SEQ ID NOS: 44
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-10-272-196-30

Query Match
Best Local Similarity 88.6%; Score 39; DB 15; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 13
US-10-612-410-18
Sequence 18, Application US/10/612,410
GENERAL INFORMATION:
APPLICANT: Bennett, Robert P.
TITLE OF INVENTION: Methods and Compositions for the Production, Identification and
FILE REFERENCE: 0942.5510003
CURRENT APPLICATION NUMBER: US/10/612,410
CURRENT FILING DATE: 2003-07-03
PRIOR APPLICATION NUMBER: 60/393,756
PRIOR FILING DATE: 2002-07-08
PRIOR APPLICATION NUMBER: 60/396,627
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/417,172
PRIOR FILING DATE: 2002-10-10
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patentin version 3.2
SEQ ID NO 18
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Strep epitope
US-10-612-410-18

Query Match
Best Local Similarity 88.6%; Score 39; DB 16; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 14
US-10-692-071-3
Sequence 3, Application US/10/692,071
GENERAL INFORMATION:
APPLICANT: Kobilka, Brian K.
APPLICANT: Gnanou, Pejman
TITLE OF INVENTION: Conformational assays to detect binding
FILE REFERENCE: STAN-213CIP
CURRENT APPLICATION NUMBER: US/10/692,071
CURRENT FILING DATE: 2003-10-22
PRIOR APPLICATION NUMBER: PCT/US02/13250
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 09/935,061
```

PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: 60/286,250
PRIOR FILING DATE: 2001-04-24
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: epitope tag peptide
US-10-692-071-3

Query Match 88.6%; Score 39; DB 16; Length 9;
Beet Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

RESULT 15
US-10-397-438A-6
Sequence 6, Application US/10397438A
Publication No. US20040191869A1
GENERAL INFORMATION:
APPLICANT: Pharmacia & Upjohn AB
TITLE OF INVENTION: Fusion vectors
FILE REFERENCE: 00126
CURRENT APPLICATION NUMBER: US/10/397,438A
CURRENT FILING DATE: 2002-11-05
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Strep-tag
US-10-397-438A-6

Query Match 88.6%; Score 39; DB 16; Length 9;
Beet Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

Search completed: March 2, 2005, 14:18:51
Job time : 25.878 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments) 65.585 Million cell updates/sec

Title: SEQ7
Perfect score: 44
Sequence: 1 rxhpqfsg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUTS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	9	1	US-08-660-626-3
2	39	88.6	9	3	US-08-828-741B-14
3	39	88.6	9	3	US-08-948-097-1
4	39	88.6	9	3	US-09-031-168-3
5	39	88.6	9	3	US-09-160-567-14
6	39	88.6	9	4	US-09-672-233-3
7	39	88.6	9	4	US-09-710-299-14
8	39	88.6	9	4	US-09-509-031-14
9	39	88.6	9	4	US-09-669-516C-3
10	39	88.6	10	1	US-08-294-386C-11
11	39	88.6	10	2	US-08-737-316A-5
12	39	88.6	10	3	US-08-897-020-6
13	39	88.6	10	3	US-08-895-707-11
14	39	88.6	10	3	US-09-350-823-6
15	39	88.6	10	4	US-09-715-805-13
16	39	88.6	10	4	US-09-692-945-9
17	39	88.6	10	4	US-09-640-041-7
18	39	88.6	10	5	PCT-US95-10224-11
19	39	88.6	12	3	US-08-218-369-8
20	39	88.6	12	4	US-09-904-599A-8
21	39	88.6	12	5	PCT-US95-03742-8
22	39	88.6	15	1	US-08-664-449-33
23	39	88.6	15	3	US-08-828-741B-11
24	39	88.6	15	3	US-09-160-567-11
25	39	88.6	15	4	US-09-710-299-11
26	39	88.6	15	4	US-09-509-031-11
27	39	88.6	17	3	US-08-828-741B-13

28	39	88.6	178	3	US-09-160-567-13	Sequence 13, Appl
29	39	88.6	178	4	US-09-710-299-13	Sequence 13, Appl
30	39	88.6	178	4	US-09-509-031-13	Sequence 13, Appl
31	39	88.6	197	3	US-08-897-020-7	Sequence 7, Appl
32	39	88.6	197	3	US-09-350-823-7	Sequence 7, Appl
33	39	88.6	198	3	US-08-828-741B-8	Sequence 8, Appl
34	39	88.6	198	3	US-09-160-567-8	Sequence 8, Appl
35	39	88.6	198	4	US-09-710-299-8	Sequence 8, Appl
36	39	88.6	198	4	US-09-509-031-8	Sequence 8, Appl
37	39	88.6	334	4	US-10-017-736C-4	Sequence 4, Appl
38	39	88.6	342	3	US-08-828-741B-6	Sequence 6, Appl
39	39	88.6	342	3	US-09-160-567-6	Sequence 6, Appl
40	39	88.6	342	4	US-09-710-299-6	Sequence 6, Appl
41	39	88.6	342	4	US-09-509-031-6	Sequence 6, Appl
42	39	88.6	386	3	US-08-895-707-7	Sequence 7, Appl
43	39	88.6	409	4	US-10-017-736C-2	Sequence 2, Appl
44	39	88.6	487	4	US-09-270-767-43055	Sequence 43055, A
45	39	88.6	495	3	US-08-828-741B-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-660-626-3
; Sequence 3, Application US/08660626
; Patent No. 5789655
; GENERAL INFORMATION:
; APPLICANT: Stanley B. Prusiner
; APPLICANT: Glenn C. Telling
; APPLICANT: Fred E. Cohen
; APPLICANT: Michael R. Scott
; TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,626
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Valera Gregg
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 07532/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-660-626-3

Query Match 88.6%; Score 39; DB 1; Length 9;
Best Local Similarity 85.7%; Pred. No. 41e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0;
Gaps 0;
QY 2 XHPQFG 8
:|||||
Db 3 RHHPFG 9

```

RESULT 2
US-08-828-741B-14
; Sequence 14, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koenigen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tatlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-828-741B-14

Query Match      88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Cy      2 XHPQFG 8
      :|||||
      :|||||
Db      3 RHPQFG 9

RESULT 3
US-08-948-097-1
; Sequence 1, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skerra, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Mutelins
; FILE REFERENCE: HUBR 1119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; CURRENT FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
; LENGTH: 9
; TYPE: PRT

```

```

; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-1

Query Match      88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Cy      2 XHPQFG 8
      :|||||
      :|||||
Db      3 RHPQFG 9

RESULT 4
US-09-031-168-3
; Sequence 3, Application US/09031168
; Patent No. 6150583
; GENERAL INFORMATION:
; APPLICANT: Stanley B. Prusiner
; APPLICANT: Glenn C. Telling
; APPLICANT: Fred E. Cohen
; APPLICANT: Michael R. Scott
; TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING
; TITLE OF INVENTION: EPTOPE-TAGGED PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Asciii
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,168
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,626
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Valeta Gregg
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 07532/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-031-168-3

Query Match      88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Cy      2 XHPQFG 8
      :|||||
      :|||||
Db      3 RHPQFG 9

RESULT 5
US-09-160-567-14
; Sequence 14, Application US/09160567

```



```
Patent No. 6326179
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Sues, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
  PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
  ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
  STREET: 400 Garden City Plaza
  CITY: Garden City
  STATE: New York
  COUNTRY: United States of America
  ZIP: 11530
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/160,567
  FILING DATE:
  CLASSIFICATION:
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/828,741
  FILING DATE:
  ATTORNEY/AGENT INFORMATION:
  NAME: DIGIGLIO, Frank S.
  REGISTRATION NUMBER: 31,346
  REFERENCE/DOCKET NUMBER: 10591
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (516) 742-4343
  TELEFAX: (516) 742-4366
  TELEX: 230 901 SANS UR
  INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 9 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
US-09-160-567-14

Query Match      88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 6
US-09-672-239-3
Sequence 3, Application US/09672239
Patent No. 6448377
GENERAL INFORMATION:
APPLICANT: Lee, Tae Weon
APPLICANT: Koblik, Brian
TITLE OF INVENTION: MODIFIED G PROTEIN SUBUNITS
FILE REFERENCE: STAN-204
CURRENT APPLICATION NUMBER: US/09/672,239
CURRENT FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-672-239-3
```

```
Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 7
US-09-710-299-14
Sequence 14, Application US/09710299
Patent No. 6521741
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Sues, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
  PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
  ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
  STREET: 400 Garden City Plaza
  CITY: Garden City
  STATE: New York
  COUNTRY: United States of America
  ZIP: 11530
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/710,299
  FILING DATE: 09-No. 6521741-2000
  CLASSIFICATION: <Unknown>
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/828,741
  FILING DATE: <Unknown>
  ATTORNEY/AGENT INFORMATION:
  NAME: DIGIGLIO, Frank S.
  REGISTRATION NUMBER: 31,346
  REFERENCE/DOCKET NUMBER: 10591
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (516) 742-4343
  TELEFAX: (516) 742-4366
  TELEX: 230 901 SANS UR
  INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 9 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-710-299-14

Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 8
US-09-509-031-14
Sequence 14, Application US/09509031
Patent No. 6590080
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
```

```

; APPLICANT: Sueess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Trentlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIPOIDS AND A METHOD OF PRODUCING SAME
; FILE REFERENCE: 13474
; CURRENT APPLICATION NUMBER: US/09/509,031
; CURRENT FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-09-509-031-14

Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFGG 8
       :|||||
DB      3 RHPQFGG 9

RESULT 9
US-09-669-516C-3
; Sequence 3, Application US/09669516C
; Patent No. 6602672
; GENERAL INFORMATION:
; APPLICANT: Prusiner, Stanley B.
; APPLICANT: Telling, Glenn C.
; APPLICANT: Cohen, Fred E.
; APPLICANT: Scott, Michael R.
; TITLE OF INVENTION: RECOMBINANT CONSTRUCT ENCODING EPTOPE
; FILE REFERENCE: UCAL-045CON
; CURRENT APPLICATION NUMBER: US/09/669,516C
; CURRENT FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 09/031,168
; PRIOR FILING DATE: 1998-02-26
; PRIOR APPLICATION NUMBER: 08/660,626
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: 08/521,992
; PRIOR FILING DATE: 1995-08-31
; PRIOR APPLICATION NUMBER: 08/509,261
; PRIOR FILING DATE: 1995-07-31
; PRIOR APPLICATION NUMBER: 08/242,188
; PRIOR FILING DATE: 1994-05-13
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strep peptide
US-09-669-516C-3

Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFGG 8
       :|||||
DB      3 RHPQFGG 9

RESULT 10
US-08-294-386C-11
; Sequence 11, Application US/08294386C
; Patent No. 5646030
```

```

; GENERAL INFORMATION:
; APPLICANT: Ray, Bryan L.
; APPLICANT: Lin, Edmund C.C.
; APPLICANT: Crea, Roberto
; TITLE OF INVENTION: Method Of Isolating Mutant Cells
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Iaplin & Kusmer
; STREET: 200 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,386C
; FILING DATE: August 23, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: SYZZ-010CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/330-1300
; TELEFAX: 617/330-1311
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
US-08-294-386C-11

Query Match      88.6%; Score 39; DB 1; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.66;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFGG 8
       :|||||
DB      4 RHPQFGG 10

RESULT 11
US-08-737-316A-5
; Sequence 5, Application US/08737316A
; Patent No. 5849576
; GENERAL INFORMATION:
; APPLICANT: SKERRA, Arne
; APPLICANT: WARDENBERG, Christina
; TITLE OF INVENTION: USE OF THE TETRACYCLINE PROMOTER FOR THE
; TITLE OF INVENTION: STRONGLY REGULATED PRODUCTION OF RECOMBINANT PROTEINS IN PR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,316A
; FILING DATE: 12-NOV-1996
```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/01862
FILING DATE: 17-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 17 598.1
FILING DATE: 19-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kiteb, Monica C.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P564-6012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-737-316A-5

Query Match
Best Local Similarity 88.6%; Score 39; DB 2; Length 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
DB 4 RHPORFG 10

RESULT 12
US-08-897-020-6
Sequence 6, Application US/08897020
Patent No. 6028176
GENERAL INFORMATION:
APPLICANT: Shanafelt, Armen; Greve, Jeffrey; Rocznak, Steven
TITLE OF INVENTION: High-affinity Interleukin-4 Molecules
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bayer Corporation, Pharmaceutical Division
STREET: 400 Morgan Lane
CITY: West Haven
STATE: CT
COUNTRY: United States of America
ZIP: 06516-4175
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS V. 6.30
SOFTWARE: Word for Windows 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,020
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P-91,242
FILING DATE: 19-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Huw R. Jones
REGISTRATION NUMBER: 33, 916
REFERENCE/DOCKET NUMBER: WH5020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 812-2317
TELEFAX: (203) 812-5492
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 10
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
DESCRIPTION: tag for streptavidin
HYPOTHETICAL: no

ANTI-SENSE: no
US-08-897-020-6

Query Match
Best Local Similarity 88.6%; Score 39; DB 3; Length 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
DB 4 RHPORFG 10

RESULT 13
US-08-895-707-11
Sequence 11, Application US/08895707
Patent No. 6077700
GENERAL INFORMATION:
APPLICANT: (Pharmacia & Upjohn, Co.)
APPLICANT: alternatively, for U.S. filing:
APPLICANT: Hollingsworth, Robert A.
APPLICANT: Sharma, Satish K.
APPLICANT: Rank, Kenneth B.
APPLICANT: Evans, David B.
TITLE OF INVENTION: Special Constructs and Complexes of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pharmacia & Upjohn Company
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Rel. #1.0, Ver. #1.25/WordPerfect 5.2+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,707
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woonton, Thomas A.
REGISTRATION NUMBER: 35,004
REFERENCE/DOCKET NUMBER: 6054
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-7914
TELEFAX: 616-833-8897
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-895-707-11

Query Match
Best Local Similarity 88.6%; Score 39; DB 3; Length 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
DB 4 RHPORFG 10

RESULT 14
US-09-350-823-6
Sequence 6, Application US/09350823

Patent No. 6313272
GENERAL INFORMATION:
APPLICANT: Shansfeldt, Armen; Greve, Jeffrey; Roczniaik, Steven
TITLE OF INVENTION: High-affinity Interleukin-4 Muteins
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSER: Bayer Corporation, Pharmaceutical Division
STREET: 400 Morgan Lane
CITY: West Haven
STATE: CT
COUNTRY: United States of America
ZIP: 06516-4175
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS v. 6.30
SOFTWARE: Word for Windows 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,823
FILING DATE: 09-Jul-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/897,020
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Huw R. Jones
REGISTRATION NUMBER: 33, 916
REFERENCE/DOCKET NUMBER: WH5020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 812-2317
TELEFAX: (203) 812-5492
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 10
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
DESCRIPTION: tag for streptavidin
HYPOTHETICAL: no
ANTI-SENSE: no
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-350-823-6
Query Match 88.6%; Score 39; DB 3; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.66;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;
QY 2 XHPQFGG 8
:|||||
Db 4 RHPQFGG 10
RESULT 15
US-09-715-805-13
Sequence 13, Application US/09715805
Patent No. 6716626
GENERAL INFORMATION:
APPLICANT: Itoh, No. 6716626uyuki
APPLICANT: Kavanagh, W. Michael
TITLE OF INVENTION: HUMAN FGF-21 GENE AND GENE EXPRESSION
PRODUCTS
FILE REFERENCE: PP-16758.001/201130.408
CURRENT APPLICATION NUMBER: US/09/715,805
CURRENT FILING DATE: 2000-11-16
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 10
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Residues which bind to paramagnetic streptavidin

OTHER INFORMATION: beads (used for purification).
US-09-715-805-13
Query Match 88.6%; Score 39; DB 4; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.66;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;
QY 2 XHPQFGG 8
:|||||
Db 4 RHPQFGG 10
Search completed: March 2, 2005, 12:25:34
Job time: 9.10569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ8
Perfect score: 42
Sequence: 1 rxhpgfek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	300	2	G75436 conserved hypothet
2	36	85.7	332	2	AF3304 oxidoreductase (EC
3	36	85.7	360	2	B71073 probable malate de
4	35	85.7	529	2	T14947 hypothetical prote
5	35	83.3	129	2	A82547 hypothetical prote
6	34	81.0	189	2	T36642 hypothetical prote
7	34	81.0	213	2	AD2533 hypothetical prote
8	34	81.0	362	2	H75131 malate dehydrogena
9	34	81.0	437	2	A31142 gelatinase, ovarian
10	34	81.0	444	2	T10487 glutamate dehydrog
11	34	81.0	543	2	A84134 acetolactate synth
12	34	81.0	627	2	T00484 hypothetical prote
13	34	81.0	1473	2	T38791 probable ferredoxi
14	34	81.0	3746	1	YGPLV3 alpha-aminoacyl-pyl-
15	34	81.0	3791	1	YGPLV8 alpha-aminoacyl-pyl-
16	33	78.6	101	2	T10856 carboxypeptidase C
17	33	78.6	211	2	C84162 hypothetical prote
18	33	78.6	310	2	C84701 hypothetical prote
19	33	78.6	335	2	AC3160 NAD binding oxidor
20	33	78.6	347	2	B85640 hypothetical prote
21	33	78.6	452	2	F84421 hypothetical prote
22	33	78.6	875	2	T50182 ubiquitin-specific
23	33	78.6	2139	2	T18296 myosin heavy chain
24	32	76.2	69	2	C82724 hypothetical prote
25	32	76.2	115	2	T13519 hypothetical prote
26	32	76.2	159	2	D75278 hypothetical prote
27	32	76.2	203	2	JQ1973 HTP9-A protein - m
28	32	76.2	293	2	E82582 DnaI protein XF223
29	32	76.2	312	2	S67052 hypothetical prote

30	32	76.2	313	2	T05139 hypothetical prote
31	32	76.2	324	2	T08307 hypothetical prote
32	32	76.2	338	2	T30538 heat shock protein
33	32	76.2	346	2	S09539 DNA ligase (ATP) (
34	32	76.2	354	2	A86843 prephenate dehydro
35	32	76.2	356	2	T38408 hypothetical prote
36	32	76.2	368	2	S75132 sensory transducti
37	32	76.2	385	2	A44102 di-N-acetylchitobi
38	32	76.2	406	2	AB1060 arginine deiminase
39	32	76.2	420	2	S71199 dnaI protein homol
40	32	76.2	434	2	H81709 N utilization subs
41	32	76.2	434	2	A71559 probable transcrip
42	32	76.2	441	2	C75076 heme biosynthesis
43	32	76.2	481	2	T14300 hypothetical prote
44	32	76.2	486	1	A35667 ty transcription a
45	32	76.2	491	2	T34499 hypothetical prote

ALIGNMENTS

```

RESULT 1
G75436
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: G75436
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Usterback, T.; Zaleski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 266, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; M01D:20036896; PMID:10567266
A:Accession: G75436
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <WHI>
A:References: UNIPROT:Q9RVC9; GB:AE001960; GB:AE000513; NID:G6458833; PIDN:AAF1067
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1100
A:Map position: 1
C:Superfamily: Escherichia coli ycea protein

Query Match          90.5%; Score 38; DB 2; Length 300;
Best Local Similarity 75.0%; Pred. No. 8;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPQFEK 8
DB      261 RAHPQFE 268

RESULT 2
AF3304
oxidoreductase (EC 1.1.1.-) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AF3304
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Marur, M.; Goldman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD33252; PMID:11756688
A:Accession: AF3304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <KUR>
A:Cross-References: UNIPROT:Q8Y1M3; GB:AB008917; PIDN:AAL51601.1; PID:G17982326; GSPDB:(
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0420
A:Map position: 1

```


R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iritaghi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: AD2533
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-213 <KUR>
 A/Cross-references: UNIPROT:Q8ZSF9; GB:AP003602; PIDN:BAW77194.1; PID:G17134636; GSPDB:G
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Gene: alx7551
 A/Genome: plasmid

Query Match 81.0%; Score 34; DB 2; Length 213;
 Best Local Similarity 71.4%; Pred. No. 33;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7
 :|||:
 Db 114 RWHPOFO 120

RESULT 8
 H75131
 malate dehydrogenase PAB1791 - *Pyrococcus abyssi* (strain Orsay)
 C/Species: *Pyrococcus abyssi*
 C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
 C/Accession: H75131
 R/anonymous, Gendecore
 submitted to the EMBL Data Library, July 1999
 A/Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru
 A/Reference number: A75001
 A/Accession: H75131
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-362 <KAW>
 A/Cross-references: UNIPROT:Q9V0D5; GB:AJ248285; GB:AL096836; NID:G5458067; PIDN:CAW4976
 A/Experimental source: strain Orsay
 C/Genetics:
 A/Gene: PAB1791
 C/Superfamily: malate dehydrogenase ylbC

Query Match 81.0%; Score 34; DB 2; Length 362;
 Best Local Similarity 62.5%; Pred. No. 59;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFE 8
 :|||:
 Db 311 RHPDPER 318

RESULT 9
 AJ1142
 gelsolin - African clawed frog (fragments)
 C/Species: *Xenopus laevis* (African clawed frog)
 C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jul-2004
 C/Accession: AJ1142; BJ1142
 R/Ankenbauer, T.; Kleinschmidt, J.A.; Vandekerckhove, J.; Franke, W.W.
 J. Cell Biol. 107, 1489-1498, 1988
 A>Title: Proteins regulating actin assembly in oogenesis and early embryogenesis of *Xen*
 A/Reference number: AJ1142; MUID:89008590; PMID:2844629
 A/Accession: AJ1142
 A/Molecule type: mRNA
 A/Residues: 21-437 <ANK1>
 A/Cross-references: UNIPROT:P14885; EMBL:X13319; NID:G64715; PIDN:CAA31694.1; PID:G64716
 A/Accession: BJ1142
 A/Molecule type: protein
 A/Residues: 1-20 <ANK2>
 C/Superfamily: gelsolin; gelsolin repeat homology
 C/Keywords: actin binding
 F.21-53/Domin: gelsolin repeat homology (fragment) <GEL1>

F.101-419/Domin: gelsolin repeat homology <GEL2>

Query Match 81.0%; Score 34; DB 2; Length 437;
 Best Local Similarity 71.4%; Pred. No. 72;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8
 :|||:
 Db 3 DHPDPER 9

RESULT 10
 T10487
 glutamate dehydrogenase (NADP) (EC 1.4.1.4) - *Prevotella ruminicola*
 C/Species: *Prevotella ruminicola*
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 C/Accession: T10487
 R/Wen, Z.T.; Morrison, M.
 submitted to the EMBL Data Library, December 1996
 A/Reference number: Z17049
 A/Accession: T10487
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-444 <WEN>
 A/Cross-references: UNIPROT:P95544; EMBL:U82240; NID:G1772844; PID:G1772845
 A/Experimental source: strain B14
 C/Genetics:
 A/Note: gdhA
 C/Superfamily: glutamate dehydrogenase (NAD(P)+)
 C/Keywords: NADP; oxidoreductase

Query Match 81.0%; Score 34; DB 2; Length 444;
 Best Local Similarity 71.4%; Pred. No. 73;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8
 :|||:
 Db 37 KHPEFEK 43

RESULT 11
 AB4134
 acetylacetate synthase large subunit BH3873 [imported] - *Bacillus halodurans* (strain C-12
 C/Species: *Bacillus halodurans*
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 16-Aug-2004
 C/Accession: AB4134
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Meeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A/Reference number: AB3650; MUID:20512582; PMID:11058132
 A/Accession: AB4134
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-543 <STO>
 A/Cross-references: UNIPROT:Q9K659; GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BAW075
 A/Experimental source: strain C-125
 C/Genetics:
 A/Gene: BH3873
 C/Superfamily: Acetylacetate synthase, large subunit/pyruvate oxidase

Query Match 81.0%; Score 34; DB 2; Length 543;
 Best Local Similarity 62.5%; Pred. No. 91;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 :|||:
 Db 470 RHPDPER 477

RESULT 12
 T00484
 hypothetical protein At2G5030 [imported] - *Arabidopsis thaliana*
 N/Alternate names: hypothetical protein F1913.26

C.Species: Arabidopsis thaliana (mouse-ear cress)
 C.Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C.Accession: T00484; G84763
 R.Roundley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
 submitted to the EMBL Data Library, April 1998
 A.Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.
 A.Reference number: Z14160
 A.Accession: T00484
 A.Status: translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-627 <ROU>
 A.Cross-references: UNIPROT:064766; EMBL:AC004238; NID:g3033373; PID:g3033399
 A.Experimental source: cultivar Columbia
 R.Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Taiton, L.;
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A.Reference number: A84420; MUID:20083487; PMID:10617197
 A.Accession: G84763
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-627 <STO>
 A.Cross-references: GB:AE002093; NID:g3033399; PID:AA12843.1; GSPDB:GN00139
 C.Genetics:
 A.Gene: F1913.26; Atg35030
 A.Map position: 2

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 627;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7
 DB 590 KSHPOFE 596

RESULT 13
 T38791
 Probable ferredoxin oxidoreductase SPAC4C5.05c - fission yeast (Schizosaccharomyces pombe)
 C.Species: Schizosaccharomyces pombe
 C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C.Accession: T38791; T37495
 R.Pearson, D.; Churcher, C.M.; Barrett, B.G.; Rajandream Wood, V.
 submitted to the EMBL Data Library, August 1997
 A.Reference number: Z21811
 A.Accession: T38791
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-1473 <PEA>
 A.Cross-references: UNIPROT:O14167; EMBL:Z98560; PID:CAA11176.1; GSPDB:GN00066; SPDB:SF
 A.Experimental source: strain 972h-; cosmid c4C5
 R.Connor, R.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, November 1997
 A.Reference number: Z21719
 A.Accession: T37495
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-755 <CON>
 A.Cross-references: EMBL:AL009197; PID:CAA15714.1; GSPDB:GN00066; SPDB:SPAC10F6.01c
 C.Genetics:
 A.Gene: SPDB:SPAC4C5.05c
 A.Map position: 1

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 1473;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFE 8
 DB 1221 RDHPOFE 1228

RESULT 14
 YGPIV3
 alpha-aminoadipyl-cysteine-lysin synthetase (EC 6.-.-.-) - Penicillium chrysogenum (str
 N/Alternate names: ACV synthetase
 C.Species: Penicillium chrysogenum
 C.Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 12-Jul-2004
 C.Accession: S13134
 R.Smith, D.J.; Earl, A.J.; Turner, G.
 EMBO J. 9, 2743-2750, 1990
 A>Title: The multifunctional peptide synthetase performing the first step of penicillin
 ic synthetases.
 A.Reference number: S13134; MUID:90360984; PMID:2118102
 A.Accession: S13134
 A.Molecule type: DNA
 A.Residues: 1-3746 <SMI>
 A.Cross-references: UNIPROT:P19787; GB:X54296; NID:g3117; PID:CAA38195.1; PID:g3118
 C.Comment: This multifunctional enzyme catalyzes the formation of alpha-aminoadipyl-cyst
 e-tying out the polymerization steps to form the tripeptide, which is the first common int
 C.Genetics:
 A.Gene: pcbAB
 C.Superfamily: non-ribosomal peptide synthetase; acetate-CoA ligase homology; acyl carri
 C.Keywords: carrier protein; cephalosporin biosynthesis; duplication; ligase; penicillin
 F1338-804/Domain: acetate-CoA ligase homology <ACLI>
 F1821-891/Domain: acyl carrier protein homology <ACPI>
 F1432-1889/Domain: acetate-CoA ligase homology <ACLI2>
 F1905-1975/Domain: acyl carrier protein homology <ACPI2>
 F12515-2977/Domain: acetate-CoA ligase homology <ACLI3>
 F12994-3062/Domain: acyl carrier protein homology <ACPI3>
 F1855,1939,3026/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match
 Best Local Similarity 81.0%; Score 34; DB 1; Length 3746;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7
 DB 2272 RTHPOFE 2278

RESULT 15
 YGPIV8
 alpha-aminoadipyl-cysteine-lysin synthetase (EC 6.-.-.-) - Penicillium chrysogenum (str
 N/Alternate names: ACV synthetase
 C.Species: Penicillium chrysogenum
 C.Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 12-Jul-2004
 C.Accession: A37886
 R.Diez, B.; Gutierrez, S.; Barredo, J.L.; van Solingen, P.; van der Voort, L.H.M.; Marti
 J. Biol. Chem. 265, 16358-16365, 1990
 A>Title: The cluster of penicillin biosynthetic genes. Identification and characterizati
 on of genes.
 A.Reference number: A37886; MUID:90375501; PMID:2129535
 A.Accession: A37886
 A.Molecule type: DNA
 A.Residues: 1-3791 <DIE>
 A.Cross-references: UNIPROT:P26046; GB:M57425; GB:U05604; NID:g169183; PID:AAA63415.1;
 C.Comment: This multifunctional enzyme catalyzes the formation of alpha-aminoadipyl-cyst
 e-tying out the polymerization steps to form the tripeptide, which is the first common int
 C.Genetics:
 A.Gene: pcbAB
 C.Superfamily: non-ribosomal peptide synthetase; acetate-CoA ligase homology; acyl carri
 C.Keywords: carrier protein; cephalosporin biosynthesis; duplication; ligase; penicillin
 F1368-834/Domain: acetate-CoA ligase homology <ACLI>
 F1851-923/Domain: acyl carrier protein homology <ACPI>
 F1464-1921/Domain: acetate-CoA ligase homology <ACLI2>
 F1937-2007/Domain: acyl carrier protein homology <ACPI2>
 F12547-3009/Domain: acetate-CoA ligase homology <ACLI3>
 F13026-3094/Domain: acyl carrier protein homology <ACPI3>
 F1885,1971,3058/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match
 Best Local Similarity 81.0%; Score 34; DB 1; Length 3791;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPOFE 7
|:|:|:|:
Db 2304 RTHPOFO 2310

Search completed: March 2, 2005, 12:28:54
Job time : 7.11382 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds

(Without alignments)
137.824 Million cell updates/sec

Title: SEQ8

Perfect score: 42

Sequence: 1 rxhpgfek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:*
1: uniprot_aprot:*
2: uniprot_cembi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	190.5	300	1 YB00 DEIRA	09rvy9 deinococcus
2	37	88.1	704	2 08CUD2	08cjd2 rattus norv
3	37	88.1	743	2 08CH90	08ch90 rattus norv
4	36	88.1	2058	2 07PVR9	07pvr9 anopheles g
5	36	85.7	209	2 06LFL4	06lfl4 plasmodium
6	36	85.7	313	2 09P127	09p127 homo sapien
7	36	85.7	318	2 06FZ38	06fz38 bartonella
8	36	85.7	318	2 06G2J1	06g2j1 bartonella
9	36	85.7	318	2 08FZ90	08fz90 brucella su
10	36	85.7	323	2 08LRT5	08lrt5 vibriophaga
11	36	85.7	332	2 08YIM3	08yim3 brucella me
12	36	85.7	360	1 MDH PYRHO	059028 pyrococcus
13	36	85.7	524	2 0934Z7	0934z7 salmonella
14	36	85.7	529	2 074YPA	074yp4 yersinia pe
15	36	85.7	529	2 06SAP7	06sap7 yersinia pe
16	36	85.7	529	2 0687A0	0687a0 yersinia pe
17	35	85.7	529	2 07ARE9	07are9 yersinia pe
18	35	83.3	129	2 09PAK0	09pak0 xyella fas
19	35	83.3	410	2 073E87	073e87 bacillus ce
20	35	83.3	496	2 08S814	08s814 bradyrhizob
21	35	83.3	10820	2 07SDK2	07sdk2 neuropeptid
22	34	81.0	70	2 06NOC1	06noc1 magnetospir
23	34	81.0	77	2 086NS2	086ns2 drosophila
24	34	81.0	91	2 08RTY3	08rty3 methanopyru
25	34	81.0	107	2 06Z315	06z315 cryza sativ
26	34	81.0	140	2 08GHJ5	08ghj5 thermus the
27	34	81.0	140	2 07ZHF1	07zhf1 thermus the
28	34	81.0	147	2 08PAR9	08par9 xanthomonas
29	34	81.0	177	2 093CR0	093cr0 shigella bo
30	34	81.0	182	2 088XG7	088xg7 lactobacill
31	34	81.0	189	2 09XBY9	09xby9 streptomyce

32	34	81.0	193	2 06ME89	06mf89 parachlamyd
33	34	81.0	207	2 06XIZ7	06xiz7 drosophila
34	34	81.0	213	2 08ZSF9	08zsf9 anabaena sp
35	34	81.0	255	2 09VQAO	09vqao drosophila
36	34	81.0	335	2 07V101	07v101 prochloroco
37	34	81.0	342	2 08TRF9	08trf9 methanosarc
38	34	81.0	348	2 06LYA4	06lya4 methanococc
39	34	81.0	362	1 MDH PYRAB	09v05 pyrococcus
40	34	81.0	376	2 07X930	07x930 prunus aviu
41	34	81.0	382	2 06XP43	06xp43 acetivibrio
42	34	81.0	384	2 07VS20	07vs20 bordetella
43	34	81.0	384	2 07WEC2	07wec2 bordetella
44	34	81.0	439	1 IDG1 DROVA	08mx40 drosophila
45	34	81.0	439	2 Q12607	Q12607 emeticella

ALIGNMENTS

RESULT 1
YB00 DEIRA STANDARD; PRT; 300 AA.
AC 09RVY9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Hypothetical UPF0176 protein DR1100.
GN OrderedLocustNames=DR1100;
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_Taxid=1299;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCBI 9279;
RX MEDLINE=20036896; PubMed=10567266; DOI=10.1126/science.286.5444.1571;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.V., Lam P., McDonald L.A., Utterback T.R., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S.L., Smith H.O., Venter J.C.,
RA Frazer C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577 (1999).
CC -1- SIMILARITY: Belongs to the UPF0176 family.
CC -1- SIMILARITY: Contains 1 rhodanese domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001960; AAF10674.1; -
CC PIR; G75436; G75436.
CC TIGR; DR1100; -
DR HAMAP; MF_00469; -; 1.
DR InterPro; IPR001792; Acylphosphatase.
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PSS0206; RHODANES3; 1.
KW Complete proteome; Hypothetical protein.
FT DOMAIN 128 222 Rhodanese.
SQ SEQUENCE 300 AA; 33238 MW; BE3D2B8C0DC43804 CRC64;

Query Match 90.5%; Score 38; DB 1; Length 300;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 DB 261 RAHPQFEE 268

RESULT 2

ID 08CUD2 PRELIMINARY; PRT; 704 AA.
 AC 08CUD2;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Guanylyl cyclase alpha 1 subunit.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura I., Suzuki N.,
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB096020; BAC24016.1; -.
 DR HSSP; P30803; IASZ.
 DR GO; GO:0004383; F:Guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR001054; G_cyclase.
 DR SMART; PF00211; Guanylate_cyc; 1.
 DR Pfam; PF00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
 KW Lyase.

SEQ SEQUENCE 704 AA; 79188 MW; F73FC97B6854E6A CRC64;
 Query Match 88.1%; Score 37; DB 2; Length 704;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPDPEK 8
 DB 698 SHPOFEK 704

RESULT 3

ID 08CH90 PRELIMINARY; PRT; 743 AA.
 AC 08CH90;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura I., Yao Y., Suzuki N.,
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB097860; BAC44887.1; -.
 DR HSSP; P30803; IASZ.
 DR GO; GO:0004383; F:Guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR001054; G_cyclase.
 DR SMART; PF00211; Guanylate_cyc; 1.
 DR Pfam; PF00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
 KW Lyase.

SEQ SEQUENCE 743 AA; 83251 MW; ACP5C53E0982813A CRC64;
 Query Match 88.1%; Score 37; DB 2; Length 743;
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPDPEK 8
 DB 737 SHPOFEK 743

RESULT 4

ID 07PUR9 PRELIMINARY; PRT; 2058 AA.
 AC 07PUR9;
 DT 01-MAR-2004 (TREMBlrel. 26, Created)
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE ENSANGP0000008445 (Fragment).
 GN Name=ENSANGP0000008445;
 OS Anopheles gambiae str. PE8T.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=PE8T;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL; AAA01008987; EAA01115.2; -.
 DR HSSP; Q24573; IC20.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR02086; Aldhyd_dehydrg.
 DR InterPro; IPR001606; ARID.
 DR Pfam; PF0388; ARID; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 FT NON TER 1
 FT NON TER 2058
 FT NON TER 2058
 SQ SEQUENCE 2058 AA; 214230 MW; BC18A6EFD0763F66 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 2058;
 Best Local Similarity 75.0%; Pred. No. 4.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 DB 922 RHHPDPEK 929

RESULT 5

ID 06FLF4 PRELIMINARY; PRT; 209 AA.
 AC 06FLF4;
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=MALEP1.55, PPF0255C;
 OS Plasmodium falciparum (Isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12368867; DOI=10.1038/nature01095;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Aktin R., Baker S., Barton A., Brooks K.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corto C.,
 RA Cronin A., Davies R., Davies P., Dear P., Dearden F., Doggett J.,

RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphrey S., Jagels K., James D., Johnson D., Kerhornou A., Knight A.,
 RA Kontofortov B., Keyes S., Larke N., Lawson D., Lennard N., Line A.,
 RA Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K.,
 RA Ormond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M.A.,
 RA Ruter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K.,
 RA Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K.,
 RA Tivey A., Unwin L., Whitehead S., Woodward J., Sulton J.E., Craig A.,
 RA Newbold C., Barrett B.G.
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungali K.,
 RA Bertman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail B.G.
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL, CR382398; CAG25222.1; -
 DR GO: GO:0003998; F:acetylphosphatase activity; IEA.
 DR InterPro: IPR001792; Acylphosphatase.
 DR Pfam: PF00708; Acylphosphatase; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 209 AA; 25373 MW; F1682AB63776B3A2 CRC64;

Query Match
 Best Local Similarity 85.7%; Score 36; DB 2; Length 209;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RXHPOFEK 8
 DB 44 KGHPEFEK 51

RESULT 6
 OP127 PRELIMINARY; PRT; 313 AA.
 AC 09P127;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE HOM-TES-85 tumor antigen (D0237H22.1).
 GN Name=du237H22.1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Testis;
 RX MEDLINE=22028805; PubMed=12032826; DOI=10.1038/9j.onc.1205481;
 RA Tureci O., Sahin U., Koslowski M., Bues B., Bell C., Ballweber P.,
 RA Zwick C., Eberle T., Zuber M., Villena-Heinsen C., Seltz G.,
 RA Pfeundschnuh M.;
 RT "A novel tumour associated leucine zipper protein targeting to sites
 of gene transcription and splicing."
 RL Oncogene 21:3879-3888(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Wilson S.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF124430; AAF28870.1; -
 DR EMBL, AL109751; CAC09922.1; -
 SQ SEQUENCE 313 AA; 35937 MW; E041911D9BA1DC8B CRC64;

Query Match
 Best Local Similarity 85.7%; Score 36; DB 2; Length 313;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 DB 182 RHHPOFEK 189

RESULT 7
 ID 06F238 PRELIMINARY; PRT; 318 AA.
 AC 06F238;
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocustNames=B009510;
 OS Bartonella quintana (Rochalimaea quintana).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bartonellaceae; Bartonella.
 OK NCBI_TaxId=803;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Toulouse;
 RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
 RA Alsmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,
 RA Canbaeck B., Eriksson A.-S., Naestlund A.K., Handley S.A., Huvet M.,
 RA La Scala B., Holmberg M., Andersson S.G.E.;
 RT "The louse-borne human pathogen Bartonella quintana is a genomic
 derivative of the zoonotic agent Bartonella henselae."
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
 DR EMBL: BX897700; CAF26428.1; -
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR InterPro: IPR000683; GFO/IDH/MOCA_N.
 DR InterPro: IPR004104; GFO/IDH/MOCA_C.
 DR Pfam: PF01408; GFO_IDH_MOCA; 1.
 DR Pfam: PF02894; GFO_IDH_MOCA_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 318 AA; 35962 MW; 638383119CD75700 CRC64;

Query Match
 Best Local Similarity 85.7%; Score 36; DB 2; Length 318;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RXHPOFEK 8
 DB 121 RHHPOFEK 128

RESULT 8
 ID 06G2J1 PRELIMINARY; PRT; 318 AA.
 AC 06G2J1;
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocustNames=BH12100;
 OS Bartonella henselae (Rochalimaea henselae).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bartonellaceae; Bartonella.
 OK NCBI_TaxId=38323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=ATCC 49882 / Houston 1;
 RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
 RA Alsmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,
 RA Canbaeck B., Eriksson A.-S., Naestlund A.K., Handley S.A., Huvet M.,
 RA La Scala B., Holmberg M., Andersson S.G.E.;
 RT "The louse-borne human pathogen Bartonella quintana is a genomic
 derivative of the zoonotic agent Bartonella henselae."
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
 DR EMBL: BX897699; CAF27993.1; -
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR InterPro: IPR000683; GFO/IDH/MOCA_N.
 DR InterPro: IPR004104; GFO_IDH_MOCA_C.
 DR Pfam: PF01408; GFO_IDH_MOCA; 1.
 DR Pfam: PF02894; GFO_IDH_MOCA_C; 1.

KM Complete proteome; Hypothetical protein.
 SQ SEQUENCE 318 AA; 35582 MW; 7AE81771FF655355 CRC64;
 Query Match 85.7%; Score 36; DB 2; Length 318;
 Best Local Similarity 75.0%; Pred. No. 85;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 :|||
 Db 121 RHPAPFEK 128

RESULT 9
 Q8FZ90 PRELIMINARY; PRT; 318 AA.
 ID Q8FZ90
 AC Q8FZ90; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Oxidoreductase, Gfo/Idh/Moca family.
 GN OrderedlocusNames=BR1602;
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122; DOI=10.1073/pnas.192319099;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Bisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.U., Umayam L.A., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J.A.,
 RA Van Aken S.E., Riedmuller S., Tettelin H., Gill S.R., White O.,
 RA Salzberg S.L., Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M.,
 RA Frazer C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL; AE014453; AAN30507.1; -.
 DR TTGR; BR1602; -.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000683; GFO/IDH/MOCA_N.
 DR InterPro; IPR004104; GFO_IDH_MOCA_C.
 DR Pfam; PF01408; GFO_IDH_MOCA_C; 1.
 DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
 KM Complete proteome.
 SQ SEQUENCE 318 AA; 35700 MW; 5367908586945BD5 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 318;
 Best Local Similarity 75.0%; Pred. No. 85;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 :|||
 Db 121 RHPAPFEK 128

RESULT 10
 Q8LTT5 PRELIMINARY; PRT; 323 AA.
 ID Q8LTT5
 AC Q8LTT5; 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE 5' to 3' exonuclease-like protein.
 OS Virioplasm VPV262.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
 OC T7-like viruses.
 OX NCBI_TaxID=194802;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=22667917; PubMed=12781722;
 RA Hardies S.C., Comeau A.M., Serwer P., Suttle C.A.;
 RT "The complete sequence of marine bacteriophage VPV262 infecting vibrio
 RT parahaemolyticus indicates that an ancestral component of a T7 viral
 RT supergroup is widespread in the marine environment."
 RL Virology 310:359-371(2003).
 DR EMBL; AY095314; AAM28378.1; -.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004527; F:exonuclease activity; IEA.
 DR InterPro; IPR008918; 5_3_exo_C.
 DR InterPro; IPR000513; Exo_N_I.
 KM Exonuclease.
 SQ SEQUENCE 323 AA; 37725 MW; 81CB297839B33951 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 323;
 Best Local Similarity 62.5%; Pred. No. 87;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 :|||
 Db 76 KKHPEFEK 83

RESULT 11
 Q8YIM3 PRELIMINARY; PRT; 332 AA.
 ID Q8YIM3
 AC Q8YIM3; 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE OXIDOREDUCTASE (EC 1.1.1.-).
 GN OrderedlocusNames=BME10420;
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16W / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688; DOI=10.1073/pnas.221575398;
 RA DelVecchio V.G., Kaputal V., Redkar R.J., Patra G., Mujter C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonowski L., Larsen H., D'Souza M., Bernal A., Mazur M., Goltsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-C.,
 RA Haselkorn R., Kyrides N.C., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009485; AAL51601.1; -.
 DR PIR; AF3304; AF3304.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR Pfam; PF01408; GFO_IDH_MOCA_C; 1.
 DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
 KM Complete proteome; Oxidoreductase.
 SQ SEQUENCE 332 AA; 37530 MW; CA57412174171CDB CRC64;

Query Match 85.7%; Score 36; DB 2; Length 332;
 Best Local Similarity 75.0%; Pred. No. 89;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
 :|||
 Db 135 RHPAPFEK 142

RESULT 12
 MDH_PYRHO STANDARD; PRT; 360 AA.
 ID MDH_PYRHO
 AC 059028;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)

```

DE Malate dehydrogenase (EC 1.1.1.37).
GN Name=mdh; OrderedLocustNames=PH1277;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxId=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=98314437; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kozugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maechi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the LDH2/MDH2 oxidoreductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AP000005; BAA30380.1; -.
DR PIR; B71073; B71073.
DR InterPro; IPR003767; Idb 2.
DR Pfam; PF02615; Idb 2; 1.
DR Complete proteome; NAD; Oxidoreductase; Tricarboxylic acid cycle.
SQ SEQUENCE 360 AA; 39751 MW; 6E9D8B16ECDBE66F CRC64;

Query Match
Best Local Similarity 85.7%; Score 36; DB 1; Length 360;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
DB 311 RKHPEFER 318

RESULT 13
Q93427 PRELIMINARY; PRT; 524 AA.
ID Q93427;
AC Q93427;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein HCM2_0069c.
GN OrderedLocustNames=HCM2_0069c;
OS Salmonella typhi.
OC Plasmid pHCM2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxId=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/25101607;
RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dow L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,

```

```

RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typh CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL513384; CAD09936.1; -.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 524 AA; 60431 MW; 99EC4CDD42F793F7 CRC64;

Query Match
Best Local Similarity 85.7%; Score 36; DB 2; Length 524;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
DB 14 RKHPEYER 21

RESULT 14
Q74YP4 PRELIMINARY; PRT; 529 AA.
ID Q74YP4;
AC Q74YP4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=PMT024;
OS Yersinia pestis.
OG Plasmid pMT1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Medievallis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.;
RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017045; AAS58659.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 529 AA; 61074 MW; 9AF9D1141768A0BB CRC64;

Query Match
Best Local Similarity 85.7%; Score 36; DB 2; Length 529;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
DB 19 RKHPEYER 26

RESULT 15
Q65AP7 PRELIMINARY; PRT; 529 AA.
ID Q65AP7;
AC Q65AP7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OG Yersinia pestis.
OC Plasmid pG8786.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC PubMed=15385458;
RA Golubov A., Neubauer H., Nolting C., Heesemann J., Rakin A.;
RT "Structural Organization of the pYra Virulence-Associated Plasmid of
RT Rhinome-Positive Yersinia pestis.";
RL Infect. Immun. 72:5613-5621(2004).
DR EMBL; AJ698720; CAG27426.1; -.

```

KW Hypothetical protein, Plasmid.
SQ SEQUENCE 529 AA; 61074 MW; 9AE89400496940BB CRC64;
Query Match 85.7%; Score 36; DB 2; Length 529;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 RXHPQFEK 8
|:|:|:
|:|:|:
Db 19 RRHPEYEK 26

Search completed: March 2, 2005, 12:44:19
Job time : 30.7236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments) 85.869 Million cell updates/sec

Title: SEQ8
Perfect score: 42
Sequence: 1 rxhpqfek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	8	AAW59212	Streptococcus
2	37	88.1	8	AAW59212	Streptococcus
3	37	88.1	8	AAW59212	Streptococcus
4	37	88.1	8	AAW59212	Streptococcus
5	37	88.1	8	AAW59212	Streptococcus
6	37	88.1	8	AAW59212	Streptococcus
7	37	88.1	8	AAW59212	Streptococcus
8	37	88.1	8	AAW59212	Streptococcus
9	37	88.1	8	AAW59212	Streptococcus
10	37	88.1	8	AAW59212	Streptococcus
11	37	88.1	8	AAW59212	Streptococcus
12	37	88.1	8	AAW59212	Streptococcus
13	37	88.1	8	AAW59212	Streptococcus
14	37	88.1	8	AAW59212	Streptococcus
15	37	88.1	8	AAW59212	Streptococcus
16	37	88.1	8	AAW59212	Streptococcus
17	37	88.1	8	AAW59212	Streptococcus
18	37	88.1	8	AAW59212	Streptococcus
19	37	88.1	8	AAW59212	Streptococcus
20	37	88.1	8	AAW59212	Streptococcus
21	37	88.1	8	AAW59212	Streptococcus
22	37	88.1	8	AAW59212	Streptococcus
23	37	88.1	8	AAW59212	Streptococcus
24	37	88.1	8	AAW59212	Streptococcus
25	37	88.1	8	AAW59212	Streptococcus

26	37	88.1	10	ADN16967	Human res
27	37	88.1	10	ADN16967	Human res
28	37	88.1	11	ADN16967	Human res
29	37	88.1	11	ADN16967	Human res
30	37	88.1	19	ABG74881	Bacterioph
31	37	88.1	19	ABG74881	Bacterioph
32	37	88.1	19	ABG74881	Bacterioph
33	37	88.1	19	ABG74881	Bacterioph
34	37	88.1	19	ABG74881	Bacterioph
35	37	88.1	19	ABG74881	Bacterioph
36	37	88.1	24	ABP60362	Streptocav
37	37	88.1	24	ABP60362	Streptocav
38	37	88.1	35	ABP60369	Streptocav
39	37	88.1	36	ABP60370	Streptocav
40	37	88.1	117	AAU97558	Synthetic
41	37	88.1	117	AAU97558	Synthetic
42	37	88.1	117	AAU97557	Synthetic
43	37	88.1	117	AAU97559	Synthetic
44	37	88.1	117	AAU97555	Synthetic
45	37	88.1	117	AAU97556	Synthetic

ALIGNMENTS

RESULT 1
ID AAW59212 standard; peptide; 8 AA.
XX AAW59212;
AC 27-AUG-1998 (first entry)
XX
DT Streptavidin tagged peptide ligand #2.
XX
DE Streptavidin tagged peptide ligand #2.
XX
KM Streptavidin; ligand; binding affinity; mutant; isolation; purification;
XX recover; immobilize.
XX
OS Synthetic.
XX
XX EP835934-A2.
XX
PN 15-APR-1998.
XX
PD 09-OCT-1997; 97EP-00117504.
XX
PF 10-OCT-1996; 96DE-01041876.
XX
PR (BIOA-) INST BIONALYTIC GMBH.
XX
PA Skerra A, Voss S;
XX
PI WPI; 1998-218868/20.
XX
DR Streptavidin mutants with higher binding affinity for peptide ligands -
XX have mutation in amino acid region 44-53, used to isolate, purify or
XX determine fusion proteins including these ligands.
XX
PS Claim 10; Page 11; 21pp; German.
XX
CC AAW59211 and AAW59212 are ligands used in a method to assay binding
XX affinity of streptavidin mutants. These mutants have a mutation within
XX the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX binding affinity than the wild-type for peptide ligands that include the
XX sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z
XX are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
XX mutants can be used to isolate, purify and determine proteins or to
XX determine/recover substances that contain streptavidin-binding groups.
XX Such compounds may also be used to immobilize fusions on microtitre
XX plates, microbeads or sensor chips
XX
SQ Sequence 8 AA;

Query Match 88.1%; Score 37; DB 2; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8
 :|||||
 DB 2 SHPQFEK 8

RESULT 2

AAB35433 standard; peptide; 8 AA.

AC AAB35433;

DT 23-MAY-2001 (first entry)

DE Eptlope peptide #3.

KW Nascent protein detection; protein analysis; aminoacylated tRNA;
 BODIPY marker; disease diagnosis.

OS Unidentified.

PN WO200114578-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000WO-US023223.

PR 25-AUG-1999; 99US-00382736.

PR 25-AUG-1999; 99US-00382950.

PA (AMBE-) AMBERGEN INC.

PI Rothschild KJ, Gite S, Olejnik J;

DR WPI; 2001-168972/17.

PT Method for detecting nascent proteins by fluorescence comprises
 misaminoacylating a tRNA molecule with a marker compound, useful for
 detecting mutations in proteins, e.g. cancer.

PS Disclosure; Page 47; 204pp; English.

CC The present invention describes a method of detecting nascent proteins
 involving aminoacylating a tRNA molecule with a 4',4-difluoro-4-bora-3a,4a
 -diaz-a-s-indacene (BODIPY) marker leading to the production of a
 misaminoacylated tRNA. This enables the detection, isolation and analysis
 of nascent proteins using UV without the usual accompanying radioactivity
 problems. It may be used to detect mutations, for example in cancer,
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer

XX Sequence 8 AA;

Query Match 88.1%; Score 37; DB 4; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8
 :|||||
 DB 2 SHPQFEK 8

RESULT 3

AAB35442 standard; peptide; 8 AA.

AC AAB35442;

DT 23-MAY-2001 (first entry)

DE Nascent protein detection method related peptide #4.

XX Nascent protein detection; protein analysis; aminoacylated tRNA;
 BODIPY marker; disease diagnosis.

OS Unidentified.

PN WO200114578-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000WO-US023223.

PR 25-AUG-1999; 99US-00382736.

PR 25-AUG-1999; 99US-00382950.

PA (AMBE-) AMBERGEN INC.

PI Rothschild KJ, Gite S, Olejnik J;

DR WPI; 2001-168972/17.

PT Method for detecting nascent proteins by fluorescence comprises
 misaminoacylating a tRNA molecule with a marker compound, useful for
 detecting mutations in proteins, e.g. cancer.

PS Example 22; Page 153; 204pp; English.

CC The present invention describes a method of detecting nascent proteins
 involving aminoacylating a tRNA molecule with a 4',4-difluoro-4-bora-3a,4a
 -diaz-a-s-indacene (BODIPY) marker leading to the production of a
 misaminoacylated tRNA. This enables the detection, isolation and analysis
 of nascent proteins using UV without the usual accompanying radioactivity
 problems. It may be used to detect mutations, for example in cancer,
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer

XX Sequence 8 AA;

Query Match 88.1%; Score 37; DB 4; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8
 :|||||
 DB 2 SHPQFEK 8

RESULT 4

AAB68616 standard; peptide; 8 AA.

AC AAB68616;

DT 27-APR-2001 (first entry)

DE Strep-Tag II sequence.

KW Protein-RNA fusion; Strep-Tag II.

OS Unidentified.

PN WO200107657-A1.

PD 01-FEB-2001.

PF 19-JUL-2000; 2000WO-US019653.

PR 27-JUL-1999; 99US-0145834P.

PA (PHYL-) PHYLLOS INC.

PI Kurz M, Lohse P, Wagner R;

DR WPI; 2001-182803/18.

CC can be used as a suppressor of immune response. (1) is useful for
 CC suppressing activation or proliferation of a cell of the immune system,
 CC suppressing IL-2 secretion by a cell, the interaction of a cell of the
 CC immune system with another cell, immunosuppressing a patient and for
 CC killing a cell expressing an antigen, human leukocyte antigen (HLA)-DR on
 CC the surface of the cell, where neither cytotoxic entities nor
 CC immunological mechanisms are needed to cause or lead to the killing. (1)
 CC (optionally linked to cytotoxic or immunogenic agent) is useful for
 CC preparing a pharmaceutical preparation for the treatment of rheumatoid
 CC arthritis, juvenile arthritis, multiple sclerosis, Grave's disease,
 CC insulin-dependent diabetes, narcolepsy, psoriasis, systemic lupus
 CC erythematosus, ankylosing spondylitis, transplant rejection, graft versus
 CC host disease, Hashimoto's disease, myasthenia gravis, pemphigus vulgaris,
 CC glomerulonephritis, thyroiditis, pancreatitis, insulinitis, primary biliary
 CC cirrhosis, irritable bowel disease and Sjogren's syndrome in humans.
 CC ABA92469 to ABA92474 and ABB57457 to ABB57590 represent sequence used in
 CC the exemplification of the present invention

XX Sequence 8 AA;

Query Match 88.1%; Score 37; DB 5; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
 :|||||
 DB 2 SHPQFEK 8

RESULT 7
 ABB77486
 ID ABB77486 standard; peptide; 8 AA.
 XX ABB77486;

XX 22-JUL-2002 (first entry)

DE AAV-helper plasmid related Strep-peptide tag SEQ ID NO 17.

KW Adeno-associated virus; AAV; glycoprotein; ITR; cap; rep; vector.

OS Synthetic.

PN WO200238782-A2.

PD 16-MAY-2002.

PF 13-NOV-2001; 2001WO-EP013125.

PR 13-NOV-2000; 2000DE-01056210.

PA (ARIM-) ARIMEDES BIOTECHNOLOGY GMBH.

PI Orberger G, Hellmuth K, Wagener C;

DR MPI; 2002-435853/46.

PT Vector system for preparing recombinant adeno-associated viral particles,
 PT used for high-level expression of heterologous therapeutic proteins in
 PT eukaryotic cells.

PS Disclosure; Page 17; 59pp; German.

CC The invention relates to a viral vector system for preparing recombinant
 CC adeno-associated virus (AAV) particles comprising: at least two plasmid
 CC vectors (AB158983 and AB158984) that include the two inverted terminal
 CC repeats (ITR) of AAV and additional sequences and plasmid vectors without
 CC ITRs but containing the rep and cap genes of AAV required for replication
 CC and packaging. The system is useful for producing recombinant AAV for
 CC production of a wide range of therapeutic glycoproteins in eukaryotic
 CC cells. The system provides efficient, large scale production of
 CC heterologous proteins in mammalian cells, without requiring an adenovirus
 CC helper. It is not toxic to host cells and does not cause lysis, so

CC produced proteins are highly pure. The present sequence is that of a
 CC peptide tag, useful to the invention

XX Sequence 8 AA;

Query Match 88.1%; Score 37; DB 5; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
 :|||||
 DB 2 SHPQFEK 8

RESULT 8
 ABG73584
 ID ABG73584 standard; protein; 8 AA.
 XX ABG73584;

DT 03-MAR-2003 (first entry)

DE StreptagII affinity peptide.

KW Non-protein L-amino acid; O-acetyl-L-serine; agrochemical production;
 KW nucleophile; O-acetyl-L-serine sulphydrylase; pharmaceutical production.

OS Synthetic.

PN EP1247869-A1.

PD 09-OCT-2002.

PF 28-MAR-2002; 2002EP-00007262.

PR 04-APR-2001; 2001DE-01016881.

PR 03-MAY-2001; 2001DE-01021515.

PA (CONE) CONSORTIUM ELEKTROCHEM IND GMBH.

PI Maier T, Gaebert C;

DR MPI; 2003-077522/08.

PT Production of non-protein L-amino acids useful for the manufacture of
 PT pharmaceuticals and agrochemicals, comprises an enzyme catalyzed reaction
 PT of O-acetyl-L-serine with a nucleophile.
 XX Example 1; Page 8; 20pp; German.

CC This invention describes a novel method for the production of non-protein
 CC L-amino acids by the reaction of O-acetyl-L-serine with a nucleophile in
 CC the presence of a catalyst comprising O-acetyl-L-serine sulphydrylase at
 CC pH 5-7.4. The method of the invention is useful for the manufacture of
 CC pharmaceuticals and agrochemicals. In contrast to the process described
 CC in DE 10046934, a high nucleophile concentration can be used which
 CC includes toxic compounds. This sequence represents an affinity peptide
 CC containing a StreptagII motif which is used in the construction of fusion
 CC genes containing Escherichia coli cyk and cym fragments

XX Sequence 8 AA;

Query Match 88.1%; Score 37; DB 6; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
 :|||||
 DB 2 SHPQFEK 8

RESULT 9
 AAE37230

ID AAE37230 standard; peptide; 8 AA.
XX
AC AAE37230;
XX
DT 07-AUG-2003 (first entry)
XX
DE Strep-tag II epitope peptide.
XX
KW Gene expression; therapy; isolation; epitope.
XX
OS Synthetic.
XX
PN WO2003038049-A2.
XX
PD 08-MAY-2003.
XX
PF 29-OCT-2002; 2002WO-US034645.
XX
PR 29-OCT-2001; 2001US-0340689P.
XX
PA (RENO-) RENOVIS INC.
XX
PI Helmut N, Serafini TA, Shyjan AW;
XX
DR WPI; 2003-430512/40.
XX
PT Isolating cell-type specific mRNAs, useful in gene expression analysis or
XX quantification in a specific cell in a heterogeneous cell mixture, by
XX isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
XX type specific manner.
XX
PS Example 2; Page 121; 136pp; English.
XX
CC The invention relates to a method for isolating mRNA from a population of
XX cells. The method involves selectively isolating ribosomes or proteins
XX that bind mRNA in a cell type specific manner and then isolating the mRNA
XX bound to the ribosomes or proteins that bind mRNA. The method is useful
XX for facilitating the analysis and quantification of gene expression in a
XX selected cell type present within a heterogeneous cell mixture. The
XX method may also be used in diagnostics or therapies for human diseases.
XX CC The present sequence is Strep-tag II epitope peptide. This sequence is
XX used to illustrate the method of the invention
XX
SQ Sequence 8 AA:
XX
Query Match 88.1%; Score 37; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQEX 8
DB 2 SHPQEX 8
XX
RESULT 10
ID ABE60361 standard; peptide; 8 AA.
XX
AC ABE60361;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin 11 tag peptide SEQ ID NO 2.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX

XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
XX protein, comprises at least two high-affinity streptavidin-binding
XX modules.
XX
PS Example 1; Page 8; 18pp; German.
XX
CC The invention relates to an isolated peptide (I) comprising at least two
XX individual modules separated by 0-50 amino acids, with each containing at
XX least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
XX streptavidin binding modules, are useful as affinity handles for
XX purification of recombinant fusion proteins (FP), also for detecting FP,
XX e.g. on protein chips or microtitre plates. The modules in (I) bind
XX strongly to streptavidin, with a co-operative effect that provides
XX stronger binding than a single tag, but are displaced by a competitor.
XX (I) does not interfere with the function of attached proteins (II) (so it
XX may not be essential to remove it); facilitates detection and has easily
XX controllable binding properties. (I) is particularly used for purifying
XX FP from dilute solution in batch formats (which use simpler apparatus
XX than column methods and result in lower loss of FP). The present sequence
XX is that of a streptavidin tag peptide disclosed with the invention
XX
SQ Sequence 8 AA:
XX

Query Match 88.1%; Score 37; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQEX 8
DB 2 SHPQEX 8
XX
RESULT 11
ID ABE60368 standard; peptide; 8 AA.
XX
AC ABE60368;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 9.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
XX protein, comprises at least two high-affinity streptavidin-binding
XX modules.
XX

PS Claim 5; Page 16; 18pp; German.
 XX
 CC The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX
 SQ Sequence 8 AA;
 XX
 Query Match 88.1%; Score 37; DB 6; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06; Indels 0; Gaps 0;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 2 XHPQFEK 8
 : |||||
 DB 2 SHPQFEK 8
 XX
 RESULT 12
 ADA09808
 ID ADA09808 standard; peptide; 8 AA.
 XX
 AC ADA09808;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Streptag epitope useful as an affinity marker.
 XX
 KW Non-radioactive marker; nascent protein detection method;
 XX cellular translation system; cell-free translation system;
 KW diethylmetheneboron difluoride dye;
 KW 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene dye; vaccine; drug;
 KW human disease screening; human disorder; protein separation;
 KW affinity marker; Streptag epitope.
 XX
 OS Synthetic.
 XX
 PN US2003092031-A1.
 XX
 PD 15-MAY-2003.
 XX
 PF 18-JUN-2002; 2002US-00174368.
 XX
 PR 25-AUG-1999; 99US-00382736.
 PR 23-AUG-2000; 2000WO-US023233.
 PR 21-JUN-2002; 2002US-00049332.
 XX
 PA (AMBE-) AMBERGEN INC.
 XX
 PI Rothschild KJ, Gite S, Olejnik J;
 XX
 DR WPI; 2003-576764/54.
 XX
 PT Detecting, analyzing or isolating nascent proteins comprises introducing
 PT a modified nucleic acid template into a cellular or cell-free translation
 PT system to generate a nascent protein having at least an N-terminal
 PT marker.
 XX
 PS Disclosure; Page 15; 76pp; English.
 XX
 CC The present invention relates to non-radioactive markers used in the
 CC detection and analysis of nascent proteins translated in cellular or cell
 CC -free translation systems. The preferred non-radioactive markers are

CC diethylmetheneboron difluoride (4,4-difluoro-4-bora-3a,4a-diaza-s-
 CC indacene) dyes. The detection method of the invention is a gel-free
 CC method that comprises introducing a modified nucleic acid template into a
 CC translation system under conditions such that a nascent protein is
 CC generated, the protein comprising at least an N-terminal marker. The
 CC method is useful in detecting, analyzing and isolating nascent proteins
 CC produced in a cell-free or cellular translation system without the use of
 CC radioactive amino acids or other radioactive labels. Compositions
 CC comprising nascent proteins translated in the presence of markers may be
 CC used as vaccines or as drugs for humans and other animals. The method and
 CC a kit containing reagents for the detection of nascent proteins may be
 CC used as a rapid means to screen humans or other animals for the presence
 CC of certain diseases or disorders. The present sequence represents an
 CC epitope that can be used as an affinity marker for protein separation.
 XX
 SQ Sequence 8 AA;
 XX
 Query Match 88.1%; Score 37; DB 7; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06; Indels 0; Gaps 0;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 2 XHPQFEK 8
 : |||||
 DB 2 SHPQFEK 8
 XX
 RESULT 13
 ADB84588
 ID ADB84588 standard; peptide; 8 AA.
 XX
 AC ADB84588;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Streptavidin conserved peptide #2.
 XX
 KW cell-free transcription system; cell-free translation system;
 KW protein synthesis; matrix; streptavidin.
 XX
 OS Escherichia coli.
 XX
 PN DE10137792-A1.
 XX
 PD 27-FEB-2003.
 XX
 PF 06-AUG-2001; 2001DE-01037792.
 XX
 PR 06-AUG-2001; 2001DE-01037792.
 XX
 PA (ERDM/) ERDMANN V.
 XX
 PI Erdmann VA, Lamla T, Stiege W;
 XX
 DR WPI; 2003-343999/33.
 XX
 PT Expressing genes in cell-free system, useful for preparation of proteins,
 PT comprises that the protein formed is removed from solution by binding to
 PT a matrix.
 XX
 PS Claim 13; Col 8; 8pp; German.
 XX
 CC This invention describes a novel method of expressing genes in a cell-
 CC free transcription and translation system which comprises using a
 CC reaction solution containing all necessary components of the
 CC transcription/translation system, amino acids, nucleotides and
 CC metabolites that supply energy and that are needed for synthesis. The
 CC proteins formed are immobilised on a matrix. The method allows simple
 CC recovery of proteins without a separate isolation step and the amount of
 CC proteins produced can be determined before a reaction is complete.
 CC Continuous removal of proteins prevents it interfering with the
 CC expression process, making possible synthesis of proteins that interact
 CC adversely with the process, so normally produced only in very low yields.
 CC The use of a matrix concentrates the proteins produced and is applicable

CC to proteins of any size. This sequence represents a highly conserved
 CC peptide from E. coli streptavidin which is used to illustrate the method
 CC of the invention.

CC Sequence 8 AA;

Query Match 88.1%; Score 37; DB 7; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQPEK 8
 :|||||
 Db 2 SHPQPEK 8

RESULT 14

ADB85500
 ID ADB85500 standard; peptide; 8 AA.

AC ADB85500;

DT 04-DEC-2003 (first entry)

DE Streptavidin tag peptide related to human aggrecanase.

KM aggrecanase; aggrecan; articular cartilage; proteoglycan;
 KM cartilage compressibility; cartilage elasticity; arthritic disease;
 KM osteoarthritis; cartilage degradation; inflammatory joint disease;
 KM aggrecanase-1; ADAMTS-4; aggrecanase-2; ADAMTS-5; chondropodondin domain;
 KM TSP domain; osteopathic; antirheumatic; cytoskeletal; antiinflammatory;
 KM antirheumatic; ophthalmological; thrombolytic; vasotropic; antimicrobial;
 KM respiratory-gen; nocotropic; neuroprotective; antiparkinsonian;
 KM immunosuppressive; aggrecanase inhibition; cancer; rheumatoid arthritis;
 KM septic arthritis; corneal ulceration; coronary thrombosis;
 KM Crohn's disease; emphysema; Alzheimer's disease; Parkinson's disease;
 KM multiple sclerosis; aortic aneurysm; streptavidin tag.

OS Synthetic.

PN WO2003066822-A2.

PD 14-AUG-2003.

PF 05-FEB-2003; 2003WO-US003554.

PR 05-FEB-2002; 2002US-0354592P.

PA (AMHP) WYETH.

PI Georgiadis K, Crawford TK, Tomkinson KM, Morris EA, Racie L;

DR WPI; 2003-731495/69.

DR N-PSDB; ADB85498.

PT New biologically-active aggrecanase protein having a deletion of all, or
 PT a portion of a TSP domain, useful for treating osteoarthritis, cancer,
 PT Parkinson's disease, coronary thrombosis, Alzheimer's disease and
 PT multiple sclerosis.

PS Example 1; Fig 18; 111pp; English.

CC This invention relates to novel truncated human aggrecanase proteins and
 CC nucleotide sequences. Aggrecan is a major extracellular component of
 CC articular cartilage. It is a proteoglycan responsible for providing
 CC cartilage with its mechanical properties of compressibility and
 CC elasticity. The loss of aggrecan has been implicated in the degradation
 CC of articular cartilage in arthritic diseases such as osteoarthritis.
 CC Aggrecanase is responsible for the cleavage of aggrecan, thereby having a
 CC role in cartilage degradation associated with osteoarthritis and
 CC inflammatory joint disease. The proteins of the current invention are
 CC truncated aggrecanase-1 (ADAMTS-4) and aggrecanase-2 (ADAMTS-5) enzymes
 CC which have at least one chondropodondin (TSP) domain deleted. These are
 CC biologically active and have greater stability and higher expression than

CC their full-length counterparts. The proteins of the invention may be of
 CC use in the development of compounds with osteopathic, antiarthritic,
 CC cytoskeletal, antiinflammatory, antirheumatic, ophthalmological,
 CC thrombolytic, vasotropic, antimicrobial, respiratory-gen, nocotropic,
 CC neuroprotective, antiparkinsonian or immunosuppressive activities through
 CC aggrecanase inhibition. The proteins of the invention may therefore be
 CC useful for the manufacture of compositions for the treatment of
 CC aggrecanase-associated conditions, such as osteoarthritis, cancer,
 CC inflammatory joint disease, rheumatoid arthritis, septic arthritis,
 CC corneal ulceration, coronary thrombosis, Crohn's disease, emphysema,
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis and aortic
 CC aneurysm. The present sequence is that of a streptavidin tag peptide
 CC linker which was used during the exemplification of the invention.

QY Sequence 8 AA;

Query Match 88.1%; Score 37; DB 7; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQPEK 8
 :|||||
 Db 2 SHPQPEK 8

RESULT 15

ADD29930
 ID ADD29930 standard; peptide; 8 AA.

AC ADD29930;

DT 15-JAN-2004 (first entry)

DE Antibody-presenting hollow protein nanoparticle-related peptide #2.

KM drug composition; hollow protein nanoparticle; cell specific antibody;
 KM cancer; virus infection; cytostatic; virucide.

OS Unidentified.

PN WO2003082330-A1.

PD 09-OCT-2003.

PF 26-MAR-2003; 2003WO-JP003694.

PR 29-MAR-2002; 2002JP-00097424.

PR 21-FEB-2003; 2003JP-00045088.

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

PI Kuroda S, Tanizawa K, Kondo A, Ueda M, Seno M, Okajima T;

DR WPI; 2003-853719/79.

PT Hollow protein nanoparticles enclosing a drug substance and presenting a
 PT tissue or cell specific antibody at their surface for site-specific
 PT therapy of cancer and other diseases.

PS Example C; SEQ ID NO 28; 96pp; Japanese.

CC The invention comprises a drug composition consisting of hollow protein
 CC nanoparticles which enclose a drug substance and which present a cell
 CC specific antibody on their surface. The drug composition is useful for
 CC cell or tissue specific treatment of cancer and virus infections. The
 CC present amino acid sequence was used in the exemplification of the
 CC invention.

QY Sequence 8 AA;

Query Match 88.1%; Score 37; DB 7; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 XHPQPEK 8
: |||||
Db 2 SHPQPEK 8

Search completed: March 2, 2005, 13:02:46
Job time : 37.0325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ8

Perfect score: 42

Sequence: 1 rxhpfek 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	36	14	US-10-026-578B-11
2	37	88.1	8	9	US-09-809-517A-9
3	37	88.1	8	9	US-09-973-145-7
4	37	88.1	8	12	US-09-813-197-8
5	37	88.1	8	13	US-10-104-218-5
6	37	88.1	8	13	US-10-208-357-9
7	37	88.1	8	14	US-10-001-934-8
8	37	88.1	8	14	US-10-026-578B-2
9	37	88.1	8	14	US-10-026-578B-9
10	37	88.1	8	14	US-10-174-368A-7
11	37	88.1	8	14	US-10-264-127-8
12	37	88.1	8	15	US-10-339-712-8
13	37	88.1	8	15	US-10-339-712-67

14	37	88.1	8	15	US-10-275-046-4	Sequence 4, Appli
15	37	88.1	8	15	US-10-425-000-76	Sequence 76, Appli
16	37	88.1	8	15	US-10-424-999-25	Sequence 25, Appli
17	37	88.1	8	15	US-10-358-283-23	Sequence 23, Appli
18	37	88.1	8	16	US-10-628-432-41	Sequence 41, Appli
19	37	88.1	8	17	US-10-494-248-17	Sequence 17, Appli
20	37	88.1	8	17	US-10-634-645-11	Sequence 11, Appli
21	37	88.1	8	17	US-10-719-523-8	Sequence 8, Appli
22	37	88.1	9	9	US-09-983-067-3	Sequence 3, Appli
23	37	88.1	10	9	US-09-809-517A-6	Sequence 6, Appli
24	37	88.1	10	15	US-10-147-211A-20	Sequence 20, Appli
25	37	88.1	11	15	US-10-354-983-29	Sequence 29, Appli
26	37	88.1	11	16	US-10-628-432-25	Sequence 25, Appli
27	37	88.1	21	9	US-09-809-517A-10	Sequence 30, Appli
28	37	88.1	22	9	US-09-809-517A-13	Sequence 33, Appli
29	37	88.1	24	9	US-09-809-517A-31	Sequence 31, Appli
30	37	88.1	24	14	US-10-026-578B-3	Sequence 3, Appli
31	37	88.1	24	14	US-10-026-578B-4	Sequence 4, Appli
32	37	88.1	25	9	US-09-809-517A-34	Sequence 34, Appli
33	37	88.1	36	14	US-10-026-578B-10	Sequence 10, Appli
34	37	88.1	117	10	US-09-977-137A-4	Sequence 4, Appli
35	37	88.1	117	10	US-09-977-137A-5	Sequence 5, Appli
36	37	88.1	117	10	US-09-977-137A-7	Sequence 7, Appli
37	37	88.1	117	10	US-09-977-137A-8	Sequence 8, Appli
38	37	88.1	117	10	US-09-977-137A-9	Sequence 9, Appli
39	37	88.1	117	10	US-09-977-137A-10	Sequence 10, Appli
40	37	88.1	117	10	US-09-977-137A-11	Sequence 11, Appli
41	37	88.1	117	10	US-09-977-137A-12	Sequence 12, Appli
42	37	88.1	118	10	US-09-977-137A-6	Sequence 6, Appli
43	37	88.1	245	17	US-10-887-228A-1	Sequence 1, Appli
44	37	88.1	246	17	US-10-887-228A-9	Sequence 9, Appli
45	37	88.1	252	17	US-10-887-228A-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-10-026-578B-11
; Sequence 11, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(28)
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11
Query Match 90.5%; Score 38; DB 14; Length 36;
Best Local Similarity 75.0%; Pred. No. 11;
```

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 XHPQFEK 8
:|||||
DB 29 QSHPOFEK 36

RESULT 2

US-09-809-517A-9
; Sequence 9, Application US/09809517A
; Patent No. US2002003473A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US2002003473A1e1 methods for displaying (poly)peptides/prote
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match 88.1%; Score 37; DB 9; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
:|||||
DB 2 SHPOFEK 8

RESULT 3

US-09-973-145-7
; Sequence 7, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-06819
; CURRENT APPLICATION NUMBER: US/09/973,145
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 09/382,950
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-973-145-7

Query Match 88.1%; Score 37; DB 9; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8

DB :|||||
2 SHPOFEK 8

RESULT 4

US-09-813-197-8
; Sequence 8, Application US/09813197
; Publication No. US2005009013A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBER-03951
; CURRENT APPLICATION NUMBER: US/09/813,197
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-813-197-8

Query Match 88.1%; Score 37; DB 12; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
:|||||
DB 2 SHPOFEK 8

RESULT 5

US-10-104-218-5
; Sequence 5, Application US/10104218
; Publication No. US20020177196A1
; GENERAL INFORMATION:
; APPLICANT: MAIER, Thomas
; APPLICANT: GABERT, Carsten
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
; FILE REFERENCE: MAIER, T. ET AL.-2
; CURRENT APPLICATION NUMBER: US/10/104,218
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US20020177196A1 101 16 881.0
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
; PRIOR FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5

Query Match 88.1%; Score 37; DB 13; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
:|||||
DB 2 SHPOFEK 8

RESULT 6

US-10-208-357-9

```
; Sequence 9, Application US/10208357
; Publication No. US20020182687A1
; GENERAL INFORMATION:
; APPLICANT: Kurz, Markus
; APPLICANT: Lohse, Peter
; APPLICANT: Wagner, Richard
; TITLE OF INVENTION: Peptide Acceptor Ligation Methods
; FILE REFERENCE: 50036/031002
; CURRENT APPLICATION NUMBER: US/10/208,357
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: US/09/619,103
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,834
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9

Query Match      88.1%; Score 37; DB 13; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 7
US-10-001-934-8
; Sequence 8, Application US/10001934
; Publication No. US20030032782A1
; GENERAL INFORMATION:
; APPLICANT: NADY, ZOLTAN
; TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
; TITLE OF INVENTION: KILLING OF CELLS INCLUDING LYMPHOID TUMOR CELLS
; FILE REFERENCE: GPCG-P01-003
; CURRENT APPLICATION NUMBER: US/10/001,934
; CURRENT FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 8
US-10-026-578B-2
; Sequence 2, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
```

```
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc.feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 9
US-10-026-578B-9
; Sequence 9, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc.feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 10
US-10-174-368A-7
; Sequence 7, Application US/10174368A
; Publication No. US20030092031A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
```

```

; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBER-07145
; CURRENT APPLICATION NUMBER: US/10/174,368A
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc_feature
; OTHER INFORMATION: Synthetic
US-10-174-368A-7

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
Db      2 SHPQEX 8

RESULT 11
US-10-264-127-8
; Sequence 8, Application US/10264127
; Publication No. US20030190643A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBER-03951
; CURRENT APPLICATION NUMBER: US/10/264,127
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: US/09/382,736B
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-264-127-8

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
Db      2 SHPQEX 8

RESULT 12
US-10-339-712-8
; Sequence 8, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy

```

```

; APPLICANT: Lim, Mark
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: AMBER-07199
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,322
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-8

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
Db      2 SHPQEX 8

RESULT 13
US-10-339-712-67
; Sequence 67, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
; APPLICANT: Lim, Mark
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: AMBER-07199
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,322
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 67
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-67

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
Db      2 SHPQEX 8

RESULT 14
US-10-275-046-4
; Sequence 4, Application US/10275046

```

```

; Publication No. US20040019187A1
; GENERAL INFORMATION:
; APPLICANT: Nagy et al.
; TITLE OF INVENTION: IMMUNOMODULATORY HUMAN MHC CLASS II ANTIGEN-BINDING POLYPEPTIDES
; FILE REFERENCE: GPCG-P01-260
; CURRENT APPLICATION NUMBER: US/10/275,046
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 00110063.5
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 4
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STREP tag II
US-10-275-046-4

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      2 XHPQFEK 8
       : |||||
       2 SHPQFEK 8

```

```

RESULT 15
US-10-425-000-76
; Sequence 76, Application US/10425000
; Publication No. US20040052777A1
; GENERAL INFORMATION:
; APPLICANT: Nebbit, Mark
; APPLICANT: Cameron, Beatrice
; APPLICANT: Blanche, Francis
; TITLE OF INVENTION: K-1 Angiotensin Polypeptides and Methods for Using Them to Inhibit
; TITLE OF INVENTION: Angiotensin
; FILE REFERENCE: ST01027-B
; CURRENT APPLICATION NUMBER: US/10/425,000
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: 10/233,675
; PRIOR FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 76
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Purification tag
US-10-425-000-76

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      2 XHPQFEK 8
       : |||||
       2 SHPQFEK 8

```

```

Search completed: March 2, 2005, 14:18:51
Job time: 24.878 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ8
Perfect score: 42
Sequence: 1 rxhpfek 8

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A COMB pep:*
2: /cgn2_6/prodata/1/1aa/5B COMB pep:*
3: /cgn2_6/prodata/1/1aa/5A COMB pep:*
4: /cgn2_6/prodata/1/1aa/5B COMB pep:*
5: /cgn2_6/prodata/1/1aa/5A COMB pep:*
6: /cgn2_6/prodata/1/1aa/5B COMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	8	US-08-948-097-2	Sequence 2, Appl
2	37	88.1	8	US-09-382-950-7	Sequence 7, Appl
3	37	88.1	8	US-09-382-7358-8	Sequence 8, Appl
4	37	88.1	8	US-09-619-103-9	Sequence 9, Appl
5	37	88.1	8	US-10-104-218-5	Sequence 9, Appl
6	37	88.1	8	US-09-809-517A-9	Sequence 9, Appl
7	37	88.1	10	US-09-809-517A-6	Sequence 9, Appl
8	37	88.1	21	US-09-809-517A-30	Sequence 30, Appl
9	37	88.1	22	US-09-809-517A-33	Sequence 33, Appl
10	37	88.1	24	US-09-809-517A-31	Sequence 31, Appl
11	37	88.1	25	US-09-809-517A-34	Sequence 34, Appl
12	37	88.1	117	US-09-977-137A-4	Sequence 5, Appl
13	37	88.1	117	US-09-977-137A-5	Sequence 5, Appl
14	37	88.1	117	US-09-977-137A-7	Sequence 8, Appl
15	37	88.1	117	US-09-977-137A-8	Sequence 9, Appl
16	37	88.1	117	US-09-977-137A-9	Sequence 10, Appl
17	37	88.1	117	US-09-977-137A-10	Sequence 11, Appl
18	37	88.1	117	US-09-977-137A-11	Sequence 12, Appl
19	37	88.1	118	US-09-977-137A-12	Sequence 6, Appl
20	37	88.1	118	US-09-977-137A-6	Sequence 71, Appl
21	34	81.0	45	US-08-630-915A-71	Sequence 71, Appl
22	34	81.0	45	US-09-879-957-71	Sequence 4700, Ap
23	34	81.0	180	US-09-134-000C-4700	Sequence 22051, A
24	34	81.0	207	US-09-248-796A-22051	Sequence 3797, Ap
25	34	81.0	213	US-09-107-532A-3797	Sequence 32857, A
26	34	81.0	234	US-09-270-767-32857	Sequence 48074, A
27	34	81.0	234	US-09-270-767-48074	

28	34	81.0	275	4	US-09-270-767-58917	Sequence 58917, A
29	34	81.0	608	4	US-09-270-767-32937	Sequence 32937, A
30	34	81.0	608	4	US-09-270-767-48154	Sequence 48154, A
31	34	81.0	712	4	US-09-248-796A-18407	Sequence 18407, A
32	34	81.0	801	4	US-09-270-767-43549	Sequence 43549, A
33	34	81.0	979	1	US-08-308-881-6	Sequence 6, Appl
34	34	81.0	979	2	US-09-058-263-6	Sequence 6, Appl
35	34	81.0	979	2	US-09-059-099-6	Sequence 6, Appl
36	34	81.0	979	3	US-09-058-264-6	Sequence 6, Appl
37	34	81.0	979	4	US-09-455-962-6	Sequence 6, Appl
38	34	81.0	979	5	PCT-US95-06530-6	Sequence 12, Appl
39	34	81.0	3666	2	US-08-222-617A-12	Sequence 27, Appl
40	34	81.0	3727	2	US-08-222-617A-27	Sequence 2, Appl
41	34	81.0	3778	2	US-08-222-617A-2	Sequence 42719, A
42	33	78.6	114	4	US-09-270-767-42719	Sequence 32254, A
43	33	78.6	169	4	US-09-252-991A-32254	Sequence 11, Appl
44	33	78.6	413	4	US-09-723-546-11	Sequence 4868, Ap
45	33	78.6	482	4	US-09-107-532A-4868	

ALIGNMENTS

```

RESULT 1
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skeira, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Muteins
; FILE REFERENCE: HUBER 1119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match      88.1%; Score 37; DB 3; Length 8;
Best local similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
      :|||||
Db      2 SHPQFEK 8

RESULT 2
US-09-382-950-7
; Sequence 7, Application US/09382950
; Patent No. 6303337
; GENERAL INFORMATION:
; APPLICANT: Rothchild, Kenneth
; APPLICANT: Gite, Sednand
; APPLICANT: Olejnik, Uerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-03879
; CURRENT APPLICATION NUMBER: US/09/382,950
; CURRENT FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial/Unknown

```

```
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: ( ) - T)
/ OTHER INFORMATION: Synthetic
US-09-382-950-7
```

```
Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPEK 8
       :|||||
Db      2 SHPQPEK 8
```

```
RESULT 3
US-09-382-736B-8
/ Sequence 8, Application US/09382736B
/ Patent No. 6306628
/ GENERAL INFORMATION:
/ APPLICANT: Rothschild, Kenneth
/ APPLICANT: Gile, Sadanand
/ APPLICANT: Olejnik, Jerry
/ TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
/ FILE REFERENCE: AMBER-03951
/ CURRENT APPLICATION NUMBER: US/09/382,736B
/ CURRENT FILING DATE: 1999-08-25
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 8
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic
US-09-382-736B-8
```

```
Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPEK 8
       :|||||
Db      2 SHPQPEK 8
```

```
RESULT 4
US-09-619-103-9
/ Sequence 9, Application US/09619103
/ Patent No. 6429300
/ GENERAL INFORMATION:
/ APPLICANT: Kurtz, Markus
/ APPLICANT: Lohse, Peter
/ APPLICANT: Wagner, Richard
/ TITLE OF INVENTION: Peptide Acceptor Ligation Methods
/ FILE REFERENCE: 50036/031002
/ CURRENT APPLICATION NUMBER: US/09/619,103
/ CURRENT FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: 60/145,834
/ PRIOR FILING DATE: 1999-07-27
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 9
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: designed sequence to act as an identifying tag
US-09-619-103-9
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPEK 8
       :|||||
Db      2 SHPQPEK 8
```

```
RESULT 5
US-10-104-218-5
/ Sequence 5, Application US/10104218
/ Patent No. 6579705
/ GENERAL INFORMATION:
/ APPLICANT: MAIER, Thomas
/ APPLICANT: GAEBERT, Carsten
/ TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
/ FILE REFERENCE: MAIER, T. ET AL.-2
/ CURRENT APPLICATION NUMBER: US/10/104,218
/ CURRENT FILING DATE: 2002-03-22
/ PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
/ PRIOR FILING DATE: 2001-04-04
/ PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
/ PRIOR FILING DATE: 2001-05-03
/ NUMBER OF SEQ ID NOS: 5
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 5
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: PEPTIDE
/ LOCATION: (1)...(8)
/ OTHER INFORMATION: StreptTagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPEK 8
       :|||||
Db      2 SHPQPEK 8
```

```
RESULT 6
US-09-809-517A-9
/ Sequence 9, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 67531361 methods for displaying (poly)peptides/proteins on b
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ PRIOR FILING DATE: 2000-02-18
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 9
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPEK 8
       :|||||
```


Db 2 SHPOFEK 8

RESULT 7

US-09-809-517A-6

; Sequence 6, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 6

; LENGTH: 10

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-6

Query Match 88.1%; Score 37; DB 4; Length 10;

Best Local Similarity 85.7%; Pred. No. 0.43;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8

Db 4 SHPOFEK 10

RESULT 8

US-09-809-517A-30

; Sequence 30, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 30

; LENGTH: 21

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-30

Query Match 88.1%; Score 37; DB 4; Length 21;

Best Local Similarity 85.7%; Pred. No. 0.97;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8

Db 15 SHPOFEK 21

RESULT 9

US-09-809-517A-33

; Sequence 33, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 33

; LENGTH: 22

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-33

Query Match 88.1%; Score 37; DB 4; Length 22;

Best Local Similarity 85.7%; Pred. No. 1;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8

Db 16 SHPOFEK 22

RESULT 10

US-09-809-517A-31

; Sequence 31, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 31

; LENGTH: 24

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-31

Query Match 88.1%; Score 37; DB 4; Length 24;

Best Local Similarity 85.7%; Pred. No. 1.1;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8

Db 18 SHPOFEK 24

RESULT 11

US-09-809-517A-34

; Sequence 34, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

```

; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```

Query Match      88.1%; Score 37; DB 4; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      19 SHPQFEK 25
```

```

RESULT 12
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4
```

```

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      111 SHPQFEK 117
```

```

RESULT 13
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
```

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5
```

```

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      111 SHPQFEK 117
```

```

RESULT 14
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7
```

```

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      111 SHPQFEK 117
```

```

RESULT 15
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match 88.1%; Score 37; DB 4; Length 117;
Best Local Similarly 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8
:|||||
Db 111 SHPOFEK 117

Search completed: March 2, 2005, 12:25:35
Job time : 10.1057 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ9

Perfect score: 42

Sequence: 1 rxhpgfer 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	360	2 B71073	probable malate de
2	37	88.1	300	2 G75436	conserved hypothet
3	37	88.1	362	2 H75131	malate dehydrogena
4	35	83.3	129	2 A82547	hypothetical prote
5	35	83.3	338	2 T10538	heat shock protein
6	35	83.3	420	2 S71199	dnau protein homol
7	34	81.0	55	2 T35092	probable small hyd
8	34	81.0	142	2 B82223	hypothetical prote
9	34	81.0	189	2 T36642	hypothetical prote
10	34	81.0	213	2 AD2533	hypothetical prote
11	34	81.0	313	2 H69297	conserved hypothet
12	34	81.0	355	2 T24938	hypothetical prote
13	34	81.0	360	2 T18140	hypothetical prote
14	34	81.0	376	2 C75580	adenine deaminase-
15	34	81.0	508	2 C82138	conserved hypothet
16	34	81.0	543	2 A84134	acetolactate synth
17	34	81.0	621	2 A71516	hypothetical prote
18	34	81.0	627	2 T00484	hypothetical prote
19	34	81.0	755	2 H72262	sensor histidine k
20	34	81.0	845	2 T07052	probable potassium
21	34	81.0	918	2 A55277	hexokinase (EC 2.7
22	34	81.0	3746	1 YGPIV3	alpha-aminoadipyl-
23	34	81.0	3791	1 YGPIV3	alpha-aminoadipyl-
24	33	78.6	105	2 T00283	hypothetical prote
25	33	78.6	310	2 C84701	hypothetical prote
26	33	78.6	332	2 AF3304	oxidoreductase (EC
27	33	78.6	347	2 B85640	hypothetical prote
28	33	78.6	360	2 F83203	probable ATP-bind
29	33	78.6	452	2 F84421	hypothetical prote

30	33	78.6	465	2 B83449	conserved hypothet
31	33	78.6	467	1 KCHUN	neutrophil collage
32	33	78.6	468	2 T32586	hypothetical prote
33	33	78.6	529	2 T14947	hypothetical prote
34	33	78.6	1577	2 T15851	hypothetical prote
35	32	76.2	69	2 C82724	hypothetical prote
36	32	76.2	73	2 S40416	hexokinase C - rat
37	32	76.2	115	2 T13519	hypothetical prote
38	32	76.2	159	2 D75278	hypothetical prote
39	32	76.2	191	2 T35792	beta-glucosidase -
40	32	76.2	203	2 JQ1973	HrpA protein - m
41	32	76.2	215	2 AC3392	maleylpyruvate iso
42	32	76.2	250	2 T34242	hypothetical prote
43	32	76.2	256	2 T10109	trypsin (EC 3.4.21
44	32	76.2	293	2 E82582	dnau protein XP223
45	32	76.2	311	2 AF2668	deacetylase [impor

ALIGNMENTS

RESULT 1
B71073
probable malate dehydrogenase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C/Accession: B71073
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A/Reference number: A71000; MUID:98344137; PMID:9679194
A/Accession: B71073
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-360 <KAW>
A/Cross-references: UNIPROT:O59028; GB:AP000005; NID:g3236132; PIDN:BAA30380.1; PID:g325
A/Experimental source: strain OT3
A/Note: This accession replaces an interim accession for a sequence replaced by GenBank
C/Genetics:
A/Gene: PH1277

Query Match 92.9%; Score 39; DB 2; Length 360;
Best Local Similarity 75.0%; Pred. No. 6.4;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPOFER 8
|:|:|:|
Db 311 RKHPFER 318

RESULT 2
G75436
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: G75436
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: AV5250; MUID:20036896; PMID:10567266
A/Accession: G75436
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-300 <WHI>
A/Cross-references: UNIPROT:Q9RCV9; GB:AE001960; GB:AE000513; NID:g6458833; PIDN:AAF1067
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR1100
A/Map position: 1
C:Superfamily: Escherichia coli ycea protein

Query Match 88.1%; Score 37; DB 2; Length 300;
 Best Local Similarity 85.7%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFE 7
 |||||
 Db 261 RAHPQFE 267

RESULT 3

H75131 malate dehydrogenase PAB1791 - *Pyrococcus abyssi* (strain Oxeay)

C/Species: *Pyrococcus abyssi*

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C/Accession: H75131

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru

A/Reference number: A75001

A/Accession: H75131

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-362 <KAM>

A/Cross-references: UNIPROT:Q9V0D5; GB:AJ248285; GB:AL096836; NID:G5458067; PIDN:CAB4976

A/Experimental source: strain Oxeay

C/Genetics:

A/Gene: PAB1791

C/Superfamily: malate dehydrogenase ylbC

Query Match 88.1%; Score 37; DB 2; Length 362;
 Best Local Similarity 75.0%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFER 8
 |||||
 Db 311 RKHPDFER 318

RESULT 4

A82547

hypothetical protein XF2515 (imported) - *Xylella fastidiosa* (strain 9a5c)

C/Species: *Xylella fastidiosa*

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C/Accession: A82547

R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A/Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A/Reference number: A82515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: A82547

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-129 <SIM>

A/Cross-references: UNIPROT:Q9PAK0; GB:AE004059; GB:AE003849; NID:G9107718; PIDN:AAFE531

A/Experimental source: strain 9a5c

R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.R.A.; Carraro, D.M.; Carver, R

de-Neto, E.; Docena, C.; El-porri, H.; Facinani, A.F.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Klafajima, J.P.; Krieger, J.B.; Kuramae, E.B.; Laigt

chado, M.A.; Madeira, A.M.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E

A/Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.U.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak

A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsuchiko, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A/Reference number: A59328

A/Contents: annotation

C/Genetics:

A/Gene: XF2515

Query Match 83.3%; Score 35; DB 2; Length 129;
 Best Local Similarity 62.5%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8
 |||||
 Db 94 RHHPQVEQ 101

RESULT 5

T30538

heat shock protein homolog dnaJ - *Trypanosoma cruzi*

C/Species: *Trypanosoma cruzi*

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

C/Accession: T30538

R/Biragaud, F.; Vedrenne, C.; Cuvilleir, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.;

Mol. Biochem. Parasitol. 94, 249-264, 1998

A/Title: Conserved organization of genes in *Trypanosomatids*.

A/Reference number: Z16580; MUID:98418771; PMID:9747975

A/Accession: T30538

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-338 <BRI>

A/Cross-references: UNIPROT:O76230; EMBL:AF031927; NID:G3452217; PID:G3452219; PIDN:AAC3

C/Genetics:

A/Gene: dnaJ

C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F:4-70/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 83.3%; Score 35; DB 2; Length 338;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8
 |||||
 Db 242 KHPQFER 249

RESULT 6

S71199

dnaJ protein homolog atj3 - *Arabidopsis thaliana*

C/Species: *Arabidopsis thaliana* (mouse-ear cress)

C/Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004

C/Accession: S71199

R/Zhou, R.; Kroczyńska, B.; Miernik, J.A.

submitted to the EMBL Data Library, March 1995

A/Description: Atj3, an *Arabidopsis thaliana* homologue of the *Escherichia coli* DnaJ.

A/Reference number: S71199

A/Accession: S71199

A/Molecule type: mRNA

A/Residues: 1-420 <ZHO>

A/Cross-references: UNIPROT:Q42530; EMBL:U22340; NID:G1872162; PIDN:ABA49030.1; PID:G727

C/Genetics:

A/Gene: atj3

C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F:14-75/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 83.3%; Score 35; DB 2; Length 420;
 Best Local Similarity 62.5%; Pred. No. 45;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8
 |||||
 Db 260 KHPQFER 267

RESULT 7

T35092

probable small hydrophilic protein - *Streptomyces coelicolor*

C/Species: *Streptomyces coelicolor*

C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C/Accession: T35092

R/Seeger, K.U.; Harris, D.; Bentley, S.D.; Parthill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999
 A:Reference number: Z21567
 A:Accession: T35092
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-55 <SEE>
 A:Cross-references: UNIPROT:Q9S286; EMBL:AL096884; PIDN:CA84410.1; GSPDB:GN00070; SCOPED
 C:Genetics:
 A:Gene: SCOEDB:SC46.32

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 55;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8
 DB 10 RXHPOAER 17

RESULT 8
 B62223
 Hypothetical protein VC1253 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C:Accession: B62223
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardon, D.; Esmolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, F.
 1, R.R.; Mekalanos, J.C.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A62035; MUID:20406833; PMID:10952301
 A:Accession: B62223
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-143 <HEI>
 A:Cross-references: UNIPROT:Q9SKJ3; GB:AE004204; GB:AE003852; NID:G9655729; PIDN:AAF9441
 A:Experimental source: serogroup O1, strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC1253
 A:Map position: 1

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 142;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8
 DB 86 KXHPQFRR 93

RESULT 9
 T36642
 Hypothetical protein SCH35.17 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: T36642
 R:Olliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1999
 A:Reference number: Z21610
 A:Accession: T36642
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-189 <OLI>
 A:Cross-references: UNIPROT:Q9X8Y9; EMBL:AL078610; PIDN:CA84410.1; GSPDB:GN00070; SCOE
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCH35.17

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 189;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7
 DB 105 KXHPQFE 111

RESULT 10
 AD2533
 Hypothetical protein alr7551 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120b
 C:Species: Nostoc sp. PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C:Accession: AD2533
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriuch
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2533
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-213 <KUR>
 A:Cross-references: UNIPROT:Q8ZSF9; GB:AP003602; PIDN:BA877194.1; PID:G17134636; GSPDB:
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr7551
 A:Genome: plasmid

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 213;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7
 DB 114 RXHPOFQ 120

RESULT 11
 H69297
 conserved hypothetical protein AF0384 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
 C:Accession: H69297
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
 ; Glock, A.; Zhou, L.; Overbeek, R.; Goeyne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
 A:Reference number: A63250; MUID:98049343; PMID:9389475
 A:Accession: H69297
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-313 <KLB>
 A:Cross-references: UNIPROT:Q29863; GB:AE001078; GB:AE000782; NID:G2689401; PIDN:AA89085

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 313;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
 DB 267 YHPOFER 273

RESULT 12
 T24938
 Hypothetical protein T15H9.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T24938
 R:Gardner, A.
 submitted to the EMBL Data Library, January 1995

A:Reference number: Z19958
A:Accession: T24938
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1355 <MIL>
A:Cross-references: UNIPROT:Q10005; EMBL:Z47356; PIDN:CAA87414.1; GSPDB:GN00020; CESP:TL
A:Experimental source: clone T15H9
C:Genetics:
A:Gene: CESP:T15H9.1
A:Map position: 2
A:introns: 42/3; 74/3; 129/3; 151/3; 208/3; 249/2; 230/3
C:Superfamily: heat shock protein dna; dna amino-terminal homology
F:4-89/Domains: dna amino-terminal homology <DNU>

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 355;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 8
DB 245 QKPRFR 252

RESULT 13
T18140
hypothetical protein A638R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18140
R:Graves, M.V.; Van Etten, J.L.
Submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T18140
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-360 <GRA>
A:Cross-references: UNIPROT:Q41120; EMBL:U42580; NID:94028896; PIDN:AAC96965.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: A638R
C:Superfamily: uncharacterized conserved protein SP0921

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 360;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
DB 233 KHPYER 239

RESULT 14
C75580
adenine deaminase-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: C75580
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <MHT>
A:Cross-references: UNIPROT:Q9KRP2; GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF1237
C:Genetics:
A:Experimental source: strain R1
A:Gene: DRA0268
A:Map position: 2

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 376;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8
DB 315 RHHPFR 322

RESULT 15
C82138
conserved hypothetical protein VC1931 [imported] - Vibrio cholerae (strain N16961 serogr
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: C82138
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: C82138
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <HEI>
A:Cross-references: UNIPROT:Q9KQR7; GB:AE004269; GB:AE003852; NID:9656466; PIDN:AAF9507
C:Genetics:
A:Gene: VC1931
A:Map position: 1
C:Superfamily: hypothetical protein bi706

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 508;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
DB 484 EHPRFR 490

Search completed: March 2, 2005, 12:28:55
Job time: 7.11382 secs


```

RP SEQUENCE FROM N.A.
RC STRAIN-OT3:
RX MEDLINE=9834437; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamuma M., Ohnuki Y.,
RA Funahashi T., Tanaka T., Kudoh K., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Ref. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the LDH2/MDH2 oxidoreductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000005; BAA30380.1; -.
DR PIR; B71073; B71073.
DR InterPro; IPR003767; 1dh_2.
DR Pfam; PF0261b; 1dh_2; 1.
DR Complete proteome; NAD: Oxidoreductase; Tricarboxylic acid cycle.
SQ SEQUENCE 360 AA; 39751 MW; 6E9DB1E6CDE6E6F CRC64;

Query Match
Best Local Similarity 92.9%; Score 39; DB 1; Length 360;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPOFE 8
Db 311 RXHPOFE 318

RESULT 3
YB00 DEIRA STANDARD; PRT; 300 AA.
AC G9RVC9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DE Hypothetical UPF0176 protein DR1100.
GN OrderedLocuNames=DR1100;
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC NCBI_TaxID=1299;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCTB 9279;
RX MEDLINE=2003696; PubMed=10567266; DOI=10.1126/science.286.5444.1571;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Mofson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Vamathevan J., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S.L., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -1- SIMILARITY: Belongs to the UPF0176 family.
CC -1- SIMILARITY: Contains 1 rhodanese domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001960; AAF10674.1; -.
DR PIR; G75436; G75436.
DR TIGR; DR1100; -.
DR HAMAP; MF_00469; -.
DR InterPro; IPR001792; Acylphosphatase.
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS0206; RHODANESF 3; 1.
KW Complete proteome; Hypothetical protein.
FT DOMAIN 128 222 Rhodanese.
SQ SEQUENCE 300 AA; 33238 MW; BE3D28BCDC43804 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 1; Length 300;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPOFE 7
Db 261 RAHPOFE 267

RESULT 4
MDH_PYRAB STANDARD; PRT; 362 AA.
AC Q9V0D5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN Name=mdh; OrderedLocuNames=PYRAB08550; ORFNames=PMB1791;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OC NCBI_TaxID=29292;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=2511545; PubMed=12622808;
RA Cohen G.N., Barde V., Flament D., Galperin M., Hellig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weisenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi."
RL Mol. Microbiol. 47:1495-1512(2003).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the LDH2/MDH2 oxidoreductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248285; CAB49769.1; -.
DR PIR; H75131; H75131.
DR InterPro; IPR003767; 1dh_2.
DR Pfam; PF0261b; 1dh_2; 1.
KW Complete proteome; NAD: Oxidoreductase; Tricarboxylic acid cycle.
SQ SEQUENCE 362 AA; 39907 MW; 1E15C2B9E6BA5012 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 1; Length 362;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches      6; 1 Conservative      1; Mismatches      1; Indels      0; Gaps      0;
QY      1 RXHPOPER 8
      :|||||
DB      311 RHPDFER 318

RESULT 5
QYX930      PRELIMINARY;      PRT;      376 AA.
ID QYX930;
AC QYX930;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
RT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
GN S-locus F-box protein 3.
OS Name=SPB3;
OS Prunus avium (Cherry);
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Rosales; Rosaceae; Amygdaloideae; Prunus.
OK NCBI_TaxId=42229;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22763733; PubMed=12881505;
RA Yamane H., Ikeda K., Ushijima K., Sassa H., Tao R.;
RT "A pollen-expressed gene for a novel protein with an F-box motif that
RT is very tightly linked to a gene for S-RNase in two species of cherry,
RT Prunus cerusue and P. avium."
RL Plant Cell Physiol. 44:764-769(2003).
DR EMBL, AB096857; BAC81148.1; -.
DR InterPro; IPR001810; F-box.
DR InterPro; IPR005527; F-box.
DR InterPro; IPR011043; Gal_oxid_central.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; F-box; 1.
DR TIGRfam; TIGR01640; F-box.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 376 AA; 43962 MW; 6076938579366587 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 376;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOPER 8
      :|||||
DB      63 HHPQFER 69

RESULT 6
QYVS20      PRELIMINARY;      PRT;      384 AA.
ID QYVS20;
AC QYVS20;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
RT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
GN Hypothetical protein.
OS OrderedLocusNames=BP0636;
OS Bordetella pertussis.
OS Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OK NCBI_TaxId=520;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parthill J., Sebalhia M., Preston A., Murphy L.D., Thomson N.R.;
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,

```

```

RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
DR EMBL, BX640412; CAE44962.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 384 AA; 39424 MW; 4A23966B197193F2 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 384;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOPER 8
      :|||||
DB      130 GHPOPER 136

RESULT 7
QYWE2      PRELIMINARY;      PRT;      384 AA.
ID QYWE2;
AC QYWE2;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
RT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
GN Hypothetical protein.
OS OrderedLocusNames=BA4712;
OC Bordetella bronchiseptica (Alcaligenes bronchiseptica).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OK NCBI_TaxId=518;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parthill J., Sebalhia M., Preston A., Murphy L.D., Thomson N.R.;
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
DR EMBL, BX640451; CAE35075.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 384 AA; 39467 MW; 779AECTC42AB9038 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 384;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOPER 8
      :|||||
DB      130 GHPOPER 136

RESULT 8
Q8XRVO      PRELIMINARY;      PRT;      557 AA.
ID Q8XRVO;
AC Q8XRVO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
RT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE PHOSPHATE SYNTHASE-6-PHOSPHATE SYNTHASE (ALPHA, ALPHA-TREHALOSE-
GN Name=R501697; OrderedLocusNames=R50731;
OS Ralstonia solanacearum (Pseudomonas solanacearum).

```


OX NCBI_TaxId=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase 1 / RSA 493;
RX MEDLINE=22608657; PubMed=12704232; DOI=10.1073/pnas.0931379100;
RA Sehnadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tetteil H., Davidsen T.M., Beaman M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.A.,
RA Khourel H.M., Lee K.H., Carty H.A., Scanlan D., Heitzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.P.,
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
RT burnetii.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
DR EMBL: AE016964; AAC090814.1; -.
DR TIGR: CBU308; -.
KW Complete proteome.
SQ SEQUENCE 221 AA; 25813 MW; 87DB00A7B91578DF CRC64;

Query Match 83.3%; Score 35; DB 2; Length 221;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8
Db 168 RGHPRFER 175

RESULT 12
06AC78 PRELIMINARY; PRT; 328 AA.
AC 06AC78;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE DnaJ protein.
GN Name=dnaJ; OrderedLocustNames=Lxxx23760;
OS Leifsonia xyli (subsp. xyli).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcales; Microbacteriaceae; Leifsonia.
OX NCBI_TaxId=59736;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CTCB07;
RX PubMed=15305603;
RA Monteiro-Vitorello C.B., Camargo L.E.A., Van Sluys M.A.,
RA Kiteajima J.P., Truffi D., do Amaral A.M., Harakava R.,
RA de Oliveira J.C.F., Wood D., de Oliveira M.C., Miyaki C.Y.,
RA Taktak M.A., da Silva A.C.R., Furlan L.R., Carraro D.M., Camarotte G.,
RA Almeida N.F., Jr., Carter H., Coutinho L.L., El-Dorry H.A.,
RA Ferro M.I.T., Gagliardi P.R., Gigliotti E., Goldman M.H.S.,
RA Goldman G.H., Kimura E.T., Ferro E.S., Kuramae E.B., Lemos E.G.M.,
RA Lemos M.V.F., Mauro S.M.Z., Machado M.A., Martino C.L., Menck C.F.,
RA Nunes L.R., Oliveira R.C., Pereira G.G., Siqueira W., de Souza A.A.,
RA Teat S.M., Zanca A.S., Simpson A.J.G., Brumley S.M., Seubal J.C.,
RT "The genome sequence of the Gram-positive sugarcane pathogen Leifsonia
RT xyli subsp. xyli.";
RL Mol. Plant Microbe Interact. 17:827-836(2004).
DR EMBL: AE016822; AAT90014.1; -.
DR GO: GO:0051082; F:unfolded protein binding; IEA.
DR GO: GO:006457; P:protein folding; IEA.
DR InterPro: IPR002939; DnaJ_C.
DR InterPro: IPR001623; DnaJ_N.
DR InterPro: IPR008971; HSP40_DnaJ_pap.
DR InterPro: IPR003095; Hsp_DnaJ.
DR Pfam: PF00226; DnaJ_1.
DR Pfam: PF01556; DnaJ_C_1.
DR PRINTS: PR00625; DnaJPROTEIN.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Chaperone; Complete proteome.
SQ SEQUENCE 328 AA; 34507 MW; 21596D5320173639 CRC64;
RT

Query Match 83.3%; Score 35; DB 2; Length 328;
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8
Db 217 RGHPRFER 224

RESULT 13
076230 PRELIMINARY; PRT; 338 AA.
ID 076230
AC 076230;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Chaperone.
GN Name=DnaJ;
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxId=5653;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL;
RX MEDLINE=98418771; PubMed=9747975; DOI=10.1016/S0166-6851(98)00080-2;
RA Brindaud F., Vedrene C., Cuivillier A., Parry D., Baltz D., Tetard E.,
RA Pays E., Venegas J., Merlin G., Baltz T.,
RT "Conserved organization of genes in trypanosomatids.";
RL Mol. Biochem. Parasitol. 94:249-264(1998).
DR EMBL: AF031927; AAC32777.1; -.
DR PIR: T30538; T30538.
DR HSP: P25685; 1HDJ.
DR GO: GO:0051082; F:unfolded protein binding; IEA.
DR GO: GO:006457; P:protein folding; IEA.
DR InterPro: IPR002939; DnaJ_C.
DR InterPro: IPR001623; DnaJ_N.
DR InterPro: IPR008971; HSP40_DnaJ_pap.
DR InterPro: IPR003095; Hsp_DnaJ.
DR Pfam: PF00226; DnaJ_1.
DR Pfam: PF01556; DnaJ_C_1.
DR PRINTS: PR00625; DnaJPROTEIN.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Chaperone.
SQ SEQUENCE 338 AA; 36535 MW; 95BA7EA791E2A19A CRC64;

Query Match 83.3%; Score 35; DB 2; Length 338;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8
Db 242 KHPHPRFER 249

RESULT 14
09B1X8 PRELIMINARY; PRT; 338 AA.
ID 09B1X8
AC 09B1X8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Co-chaperone protein.
GN Name=DnaJ-like;
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxId=5653;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21570179; PubMed=11551903; DOI=10.1074/jbc.M102427200;
RA Salmon D., Monteiro-Lomeij M., Goldenberg S.,
RT "A DnaJ-like protein homologous to the yeast co-chaperone Sisl (Tcd9p)

RT is involved in initiation of translation in *Trypanosoma cruzi*."
 RL J. Biol. Chem. 276:43970-43979(2001).
 DR EMBL; AF345335; AK19734.1; -.
 DR HSPB, P25685; IHDJ.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:006457; P:protein folding; IEA.
 DR InterPro; IPR002939; DnaJ_C.
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR008971; HSP40_DnaJ_pep.
 DR InterPro; IPR003095; Hsp_DnaJ.
 DR Pfam; PF00226; DnaJ_1.
 DR Pfam; PF01556; DnaJ_C_1.
 DR PRINTS; PR00625; DnaJPROTEIN.
 DR SMART; SM00271; DnaJ_1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 KW Chapterone.
 SQ SEQUENCE 338 AA; 36561 MW; 90A9E4E04508E655 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 338;
 Best Local Similarity 62.5%; Pred. No. 1.4e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8
 ::|||::|||
 Db 242 KHPHPR 249

RESULT 15

Q42530 PRELIMINARY; PRT; 420 AA.
 AC Q42530;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE DnaJ homolog.
 GN Name=atj;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid1; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20027448; PubMed=10557255;
 RX Zhou R., Kroczyńska B., Mierzyk J.A.,
 RT "AtCJ3 (Accession No. U22340), an Arabidopsis thaliana J-Protein
 RL Homologous to Saccharomyces cerevisiae YDulp. (PGR99-162).";
 RL Plant Physiol. 121:1053-1054(1999).
 DR EMBL; U22340; AAB49030.1; -.
 DR PIR; S71199; S71199.
 DR HSPB, P25685; IHDJ.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:006457; P:protein folding; IEA.
 DR InterPro; IPR002939; DnaJ_C.
 DR InterPro; IPR001305; DnaJ_CXXCXG.
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR008971; HSP40_DnaJ_pep.
 DR InterPro; IPR003095; Hsp_DnaJ.
 DR Pfam; PF00226; DnaJ_1.
 DR Pfam; PF01556; DnaJ_C_1.
 DR Pfam; PF00684; DnaJ_CXXCXG; 1.
 DR PRINTS; PR00625; DnaJPROTEIN.
 DR SMART; SM00271; DnaJ_1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 KW Chapterone; Repeat.
 SQ SEQUENCE 420 AA; 46444 MW; C4C12848F61AD445 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 420;
 Best Local Similarity 62.5%; Pred. No. 1.8e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8
 ::|||::|||
 Db 260 KHPHPR 267

Search completed: March 2, 2005, 12:44:20
 Job time : 30.7236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ9

Perfect score: 42

Sequence: 1 rxhpqfer 8

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	231	7	ADB65422
2	39	92.9	313	4	AAB75608
3	39	92.9	359	8	AD841924
4	39	92.9	360	8	ADN18635
5	37	88.1	11	2	AAR52694
6	37	88.1	11	2	AAR52693
7	37	88.1	12	2	AAR52697
8	37	88.1	362	8	ADS43180
9	36	85.7	482	7	ADC95241
10	36	85.7	619	6	ABU19450
11	35	83.3	418	6	ABP80316
12	35	83.3	749	4	ABR62997
13	35	83.3	2747	7	AD886071
14	34	81.0	8	2	AAR59212
15	34	81.0	8	4	AAB35433
16	34	81.0	8	4	AAB35442
17	34	81.0	8	4	AAB68616
18	34	81.0	8	5	AAO19065
19	34	81.0	8	5	ABR57464
20	34	81.0	8	5	ABR77486
21	34	81.0	8	6	ABG73584
22	34	81.0	8	6	AAE37230
23	34	81.0	8	6	ABP60361
24	34	81.0	8	6	ABP60368
25	34	81.0	8	7	ADA09808

ALIGNMENTS

26	34	81.0	8	7	ADB84588	ADB84588	Streptavi
27	34	81.0	8	7	ADB85500	ADB85500	Streptavi
28	34	81.0	8	7	ADD29930	ADD29930	Antibody-
29	34	81.0	8	8	AD006951	AD006951	Streptag
30	34	81.0	8	8	ADP90832	ADP90832	Protein/p
31	34	81.0	8	8	AD520244	AD520244	Strept tag
32	34	81.0	8	8	ADR72509	ADR72509	Dehydroge
33	34	81.0	9	5	ABG31054	ABG31054	Peptide p
34	34	81.0	10	4	AA979662	AA979662	Influenza
35	34	81.0	10	4	AA999027	AA999027	Streptavi
36	34	81.0	10	5	AA080475	AA080475	Peptide S
37	34	81.0	10	6	ABP56623	ABP56623	C-termina
38	34	81.0	10	8	ADF11064	ADF11064	Strept-tag
39	34	81.0	10	8	ADN16967	ADN16967	Human res
40	34	81.0	10	8	AD026489	AD026489	Strept-tag
41	34	81.0	11	6	AAE38373	AAE38373	Epitope t
42	34	81.0	11	8	AD520228	AD520228	Synthetic
43	34	81.0	13	2	AAE78379	AAE78379	Streptavi
44	34	81.0	19	6	ABG74881	ABG74881	Bacteriop
45	34	81.0	19	6	ABG74882	ABG74882	Bacteriop

RESULT 1

ADB65422 standard; protein, 231 AA.

ADB65422;

04-DEC-2003 (first entry)

Human protein encoded by clone TESIT120180600.

Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;

cell regeneration; membrane protein; signal transduction-related protein;

transcription-related protein; osteoporosis; neurological disease;

cancer; tumour.

Homo sapiens.

EP1308459-A2.

07-MAY-2003.

28-MAR-2002; 2002EP-00007401.

05-NOV-2001; 2001JP-00379298.

25-JAN-2002; 2002US-00350978.

(HELI-) HELIX RES INST.

(REAS-) RES ASSOC BIOTECHNOLOGY.

Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

Yamanoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;

WPI; 2003-450961/43.

N-PSDB; ADB63452.

New polynucleotides and polypeptides, useful for developing a diagnostic

marker or medicines for regulation of their expression and activity, or

as targets of gene therapy.

Claim 1, Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected

from 1970 Fully defined nucleotide sequences which encode novel

polypeptides. Also claimed is a polypeptide encoded by the polynucleotide

or its partial peptide, an antibody binding to the polypeptide or peptide

of the polynucleotide, immunologically assaying the polypeptide or

peptide of the polynucleotide by contacting the polypeptide or peptide

CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotide and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumors). The CDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a protein of the invention. Note: Some of the
CC sequence data for this patent is not represented in the printed
CC European Patent Office.

SQ Sequence 231 AA;

Query Match 92.9%; Score 39; DB 7; Length 231;

Best Local Similarity 75.0%; Pred. No. 41;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOPER 8
|:|||||
DB 100 RHHPOPER 107

RESULT 2

ID AAB75608 standard; protein; 313 AA.

XX AAB75608;

DT 10-APR-2001 (first entry)

XX Human cancer associated antigen precursor HOM-TES-85 SEQ ID NO:11.

XX Human; cancer associated antigen precursor; cancer associated antigen;

XX seminoma; HLA; human leukocyte antigen; cytostatic; gene therapy;

XX vaccine; cancer.

XX Homo sapiens.

XX WO200100874-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-US017207.

XX 30-JUN-1999; 99US-00346498.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Sahin U, Tureci O, Firenzschnuh M;

XX WPI; 2001-112465/12.

XX N-PSDB; AAF26957.

XX Diagnosing a disorder characterized by expression of a human cancer

XX with a nucleic acid molecule encoding the antigen precursor.

XX Example 2; Page 97; 126pp; English.

CC The present invention describes a method for diagnosing a disorder
CC characterized by expression of a human cancer associated antigen (CAA)
CC precursor (I) coded by a NA Group 1 nucleic acid molecule (NI) comprising
CC contacting the biological sample with an agent (A) that specifically
CC binds to NI, (I) or its fragment, complexed with an human leukocyte
CC antigen (HLA) molecule and determining the interaction between the agent

CC and NI or (I). (I) has cytostatic activity and can be used in gene
CC therapy and vaccine production. The method can be used for treating a
CC subject with a condition characterized by expression of (I) in cells of a
CC subject. AAB75607 and AAB75608 represent proteins from human cancer
CC associated antigen precursors, and AAB75609 to AAB75802 represent HLA
CC class I binding motifs in human cancer associated antigen precursors
CC given in the exemplification of the present invention

SQ Sequence 313 AA;

Query Match 92.9%; Score 39; DB 4; Length 313;

Best Local Similarity 75.0%; Pred. No. 57;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOPER 8
|:|||||
DB 182 RHHPOPER 189

RESULT 3

ID ADS41924 standard; protein; 359 AA.

XX ADS41924;

DT 02-DEC-2004 (first entry)

XX Bacterial polypeptide #20354.

XX Recombinant DNA construct; transformed plant; improved plant property;

XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

XX pathogen tolerance; pest tolerance; plant disease resistance;

XX cell cycle pathway modification; plant growth regulator;

XX homologous recombination; seed oil yield; protein yield; carbohydrate;

XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

XX Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide

XX for expression of a polynucleotide encoding a polypeptide from a

XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 20354; 122pp; English.

CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.

CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomanan
CC production, this sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
SQ Sequence 359 AA;
Query Match 92.9%; Score 39; DB 8; Length 359;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RXHPOFER 8
Db 311 RKHPEFER 318
RESULT 4
ADN18635 ADN18635 standard; protein; 360 AA.
XX
XX ADN18635; 1
XX
XX 02-DEC-2004 (first entry)
DT
XX Bacterial polypeptide #1288.
DE
XX Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;
XX cell cycle pathway modification; plant growth regulator;
XX homologous recombination; seed oil yield; protein yield; carbohydrate;
XX nitrogen; phosphorus; photosynthesis; lignin; galactomanan;
XX bacterial polypeptide.
XX
XX Bacteria.
OS
XX US2003233675-A1.
PN
XX 18-DEC-2003.
PD
XX 20-FEB-2003; 2003US-00369493.
PF
XX 21-FEB-2002; 2002US-0360039P.
PR
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS.
PI
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 1288; 122bp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a

CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomanan
CC production, this sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
SQ Sequence 360 AA;
Query Match 92.9%; Score 39; DB 8; Length 360;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RXHPOFER 8
Db 311 RKHPEFER 318
RESULT 5
AAR52694 AAR52694 standard; protein; 11 AA.
ID
XX AAR52694;
AC
XX 10-JAN-1995 (first entry)
XX
XX PASK46-p14KH encoded C-terminal streptavidin-binding sequence.
DE
XX Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
XX heavy chain variable region; affinity chromatography; purification;
XX peptide tag.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Peptide 4..11
FT /note= "streptavidin-binding octapeptide fused to C-terminus of VH chain"
FT
XX
XX GB2272698-A.
PN
XX 25-MAY-1994.
PD
XX 01-NOV-1993; 93GB-00022501.
PF
XX 03-NOV-1992; 92DE-04237113.
PR
XX (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.
XX
XX Skerra A, Schmidt T;
PI
XX WPI: 1994-153484/19.
XX
XX N-PSDB; AAO62669.
XX
XX New fusion peptide(s) - have easily controlled binding properties and are
XX capable of binding to streptavidin.
XX
XX Disclosure; Page 11; 53pp; English.
PS

XX Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv
 CC fragment in E.coli) were produced which encode 4 different peptides at
 CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv
 CC fragment. The peptides fused to the C-terminus are all examples of
 CC streptavidin-binding peptides corresponding to a generic formula (see
 CC AAR52698). The peptides do not interfere with the protein function but
 CC facilitate purification by conferring streptavidin-binding properties on
 CC the fusion protein

SQ Sequence 11 AA;

Query Match 88.1%; Score 37; DB 2; Length 11;
 Best Local Similarity 85.7%; Pred. No. 3.7;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8
 :|||||
 DB 5 LHPQFER 11

RESULT 6

AAR52693 ID AAR52693 standard; protein; 11 AA.

AC AAR52693;

DT 10-JAN-1995 (first entry)

DE pASK46-p11XH encoded C-terminal streptavidin-binding sequence.

KM Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
 KM heavy chain variable region; affinity chromatography; purification;
 KM peptide tag.

OS Synthetic.

Key Location/Qualifiers
 FH 4..11
 FT /note= "streptavidin-binding octapeptide fused to C-
 FT terminus of VH chain"

GB2272698-A.

25-MAY-1994.

01-NOV-1993; 93GB-00022501.

03-NOV-1992; 92DE-04237113.

(BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

Skerra A, Schmidt T;

WPI; 1994-153484/19.

N-PSDB; AAQ62668.

New fusion peptide(s) - have easily controlled binding properties and are
 capable of binding to streptavidin.

Disclousre; Page 11; 53pp; English.

Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv
 fragment in E.coli) were produced which encode 4 different peptides at
 the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv
 fragment. The peptides fused to the C-terminus are all examples of
 streptavidin-binding peptides corresponding to a generic formula (see
 AAR52698). The peptides do not interfere with the protein function but
 facilitate purification by conferring streptavidin-binding properties on
 the fusion protein

Sequence 11 AA;

Query Match 88.1%; Score 37; DB 2; Length 11;
 Best Local Similarity 85.7%; Pred. No. 3.7;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8
 :|||||
 DB 5 RHPQFER 11

RESULT 7

AAR52697 ID AAR52697 standard; protein; 12 AA.

AC AAR52697;

DT 10-JAN-1995 (first entry)

DE pASK46-p11XL encoded C-terminal streptavidin-binding sequence.

KM Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
 KM light chain variable region; affinity chromatography; purification;
 KM peptide tag.

OS Synthetic.

Key Location/Qualifiers
 FH 5..12
 FT /note= "streptavidin-binding octapeptide fused to C-
 FT terminus of VL chain"

GB2272698-A.

25-MAY-1994.

01-NOV-1993; 93GB-00022501.

03-NOV-1992; 92DE-04237113.

(BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

Skerra A, Schmidt T;

WPI; 1994-153484/19.

N-PSDB; AAQ62672.

New fusion peptide(s) - have easily controlled binding properties and are
 capable of binding to streptavidin.

Disclousre; Page 11; 53pp; English.

Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv
 fragment in E.coli) were produced which encode 4 different peptides at
 the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv
 fragment. The peptides fused to the C-terminus are all examples of
 streptavidin-binding peptides corresponding to a generic formula (see
 AAR52698). The peptides do not interfere with the protein function but
 facilitate purification by conferring streptavidin-binding properties on
 the fusion protein

Sequence 12 AA;

Query Match 88.1%; Score 37; DB 2; Length 12;
 Best Local Similarity 85.7%; Pred. No. 4;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8
 :|||||
 DB 6 RHPQFER 12

RESULT 8

ADS43180 ID ADS43180 standard; protein; 362 AA.

```

XX AC ADS43180;
XX DT 02-DEC-2004 (first entry)
XX DE Bacterial polypeptide #21610.
XX KM Recombinant DNA construct; transformed plant; improved plant property;
XX KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX KM pathogen tolerance; pest tolerance; plant disease resistance;
XX KM cell cycle pathway modification; plant growth regulator;
XX KM homologous recombination; seed oil yield; protein yield; carbohydrate;
XX KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX KM bacterial polypeptide.
XX OS Bacteria.
XX PN US2003233675-A1.
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.
XX PA (CAOY/) CAO Y.
XX PA (HINK/) HINKLE G J.
XX PA (SLAT/) SLATER S C.
XX PA (CHEN/) CHEN X.
XX PA (GOLD/) GOLDMAN B S.
XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX DR WPI; 2004-061375/06.
XX PT New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX PS Claim 1; SEQ ID NO 21610; 122pp; English.
XX CC The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned to
XX CC provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transformed plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX CC improved plant properties, e.g. improved cold, heat or drought tolerance,
XX CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX CC increased resistance to plant disease, better growth rate by modification
XX CC of the cell cycle pathway with plant growth regulators, increased rate of
XX CC homologous recombination, modified seed oil or protein yield and/or
XX CC content, improved yield by modification of carbohydrate, nitrogen or
XX CC phosphorus use and/or uptake, by modification of photosynthesis or by
XX CC providing improved plant growth and development under at least one stress
XX CC condition, improved lignin production or improved galactomannan
XX CC production. This sequence represents a bacterial polypeptide used in the
XX CC scope of the invention. Note: The sequence data for this patent did not
XX CC form part of the printed specification but was obtained in electronic
XX CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 362 AA;
XX
Query Match      88.1%; Score 37; DB 8; Length 362;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 RXHPQFER 8

```

```

DB          311 RXHPQFER 318
          ||:|:|
RESULT 9
ADCS2241
ID ADC95241 standard; protein; 482 AA.
XX AC ADC95241;
XX DT 01-JAN-2004 (first entry)
XX DE E. faecium protein sequence SEQ ID 4868.
XX KM Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
XX KM abdominal-pelvic infection.
XX OS Enterococcus faecium.
XX PN US6583275-B1.
XX PD 24-JUN-2003.
XX PF 30-JUN-1998; 98US-00107532.
XX PR 02-JUL-1997; 97US-0051571P.
XX PR 14-MAY-1998; 98US-0085598P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2003-799836/75.
XX DR N-PSDB; ADC91587.
XX PT New isolated nucleic acid derived from Enterococcus faecium encoding an
XX PT Enterococcus faecium polypeptide useful for detection, prevention and
XX PT treatment of a pathological condition resulting from a bacterial
XX PT infection.
XX PS Example 1; SEQ ID NO 4868; 243pp; English.
XX CC The invention relates to an isolated nucleic acid derived from
XX CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
XX CC one of 10 fully defined sequences given in the (or comprising 40
XX CC sequential nucleotides chosen from any of the nucleic acids, its
XX CC complement or sequences hybridising to it). Also included are a
XX CC recombinant vector comprising the nucleic acid operably linked to
XX CC transcription regulatory element, a cell comprising the vector and a
XX CC single-stranded probe comprising the nucleic acid. The nucleic acids are
XX CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
XX CC The nucleic acids is useful for diagnosing pathological conditions
XX CC resulting from E. faecium bacterial infection (e.g. urinary tract
XX CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
XX CC infection) and for screening drugs such as agonists and antagonists. The
XX CC nucleic acid is useful for recombinant production of Candida albicans -
XX CC derived peptides or antisense polypeptides. Pharmaceutical compositions
XX CC and vaccines containing the nucleic acid are useful for preventing or
XX CC treating Enterococcus faecium infections. The present sequence represents
XX CC one if the disclosed E. faecium proteins.
XX SQ Sequence 482 AA;
XX
Query Match      85.7%; Score 36; DB 7; Length 482;
Best Local Similarity 62.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 RXHPQFER 8
DB          409 KXHPQFER 416
          ::|:|:|
RESULT 10

```

ABU19450
ID ABU19450 standard; protein; 619 AA.
XX
AC ABU19450;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #4977.
XX
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Borrelia cepacia.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
FI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JM;
FI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
DR N-PSDB; ACP23320.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 47374; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation; (7) identifying a gene in an operon required for
XX proliferation; (8) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (9)
XX identifying a gene required for cellular proliferation or its gene product
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collecting the
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 619 AA;

Query Match 85.7%; Score 36; DB 6; Length 619;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 RXHPQFER 8
DB 339 RHFOIER 346
RESULT 11
ID ABP80316 standard; protein; 418 AA.
XX
AC ABP80316;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 7162.
XX
KM Antibacterial; infection; vaccine; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB002069.
XX
PR 12-FEB-2001; 2001GB-00003424.
XX
PA (CHIR-) CHIRON SPA.
XX
FI Fontana MR, Piza M, Maignani V, Monaci E;
XX
XX WPI; 2003-058415/05.
XX
DR N-PSDB; AB241286.
XX
PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection.
XX
PS Disclosure; Page 708; 815bp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX
XX Sequence 418 AA;
SQ
Query Match 83.3%; Score 35; DB 6; Length 418;
Best Local Similarity 62.5%; Pred. No. 4.3e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 RXHPQFER 8
DB 16 RKHPKDR 23
RESULT 12
ID ABP82997 standard; protein; 749 AA.
XX
AC ABP82997;
XX
DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 15783.
XX
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 XX
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL07100.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 15783; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (AAB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 749 AA;
 Query Match 83.3%; Score 35; DB 4; Length 749;
 Best Local Similarity 75.0%; Pred. No. 8.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RXHPOPER 8
 DB 546 RSHPOFVR 553
 RESULT 13
 ADE86071
 ID ADE86071 standard; protein; 2747 AA.
 XX
 AC ADE86071;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Streptomyces hygroscopicus non-ribosomal peptide synthetase mppA.
 XX
 KW Peptide synthetase; enzyme; antibiotic; mppA; AC98.
 XX
 OS Streptomyces hygroscopicus.
 XX
 PN WO2003082909-A1.
 XX
 PD 09-OCT-2003.
 XX
 PF 28-MAR-2003; 2003WO-US009746.
 XX
 PR 29-MAR-2002; 2002US-0368713P.
 XX
 PA (AMHP) WYETH.
 XX

PI Hucul JA, Macgarvey N, Greenstein M;
 XX
 DR WPI; 2003-812536/76.
 XX
 DR N-PSDB; ADE86070.
 XX
 PT New non-ribosomal peptide synthetase, useful for preparing antibacterial
 PT peptides, derived from Streptomyces, also related nucleic acid and
 PT modified antibiotics.
 XX
 PS Claim 10; SEQ ID NO 2; 163bp; English.
 XX
 CC The present sequence is the protein sequence of mppA, a component of the
 CC non-ribosomal peptide synthetase (NRPS) from Streptomyces hygroscopicus
 CC NS17 (NRRL 30439) that is responsible for the production of cyclic
 CC peptide antibiotic AC98. The NRPS complex exists as 2 separate
 CC components, mppA and mppB ADE86072, both involved in the synthesis of the
 CC core AC98. MppA is composed of 3 minimal modules, where each module is
 CC comprised of an adenylation, thiolation and condensation domain. MppA
 CC conjugates a serine to a glycine to produce a peptide. The peptide is
 CC then conjugated (through the glycine) to a phenylalanine. Each amino acid
 CC is incorporated into the peptide chain by a unique module. After addition
 CC of the phenylalanine, the peptide chain is then transferred to the mppB
 CC component. MppB is about 295 kDa. The invention provides a method of
 CC producing NRPS using transformed host cells. It also provides a method of
 CC producing a cyclic peptide antibiotic, such as AC98, using the NRPS
 CC comprised of mppA and mppB. The antibiotic is preferably a
 CC lipoglycopeptide antibiotic with activity against Gram-positive
 CC pathogens. Also provided are methods of modifying NRPS, e.g. by
 CC replacement of an adenylation domain, to produce an antibiotic having a
 CC modified peptide core, and a method for evaluating the structural regions
 CC of the modified peptide.
 XX
 SQ Sequence 2747 AA;
 Query Match 83.3%; Score 35; DB 7; Length 2747;
 Best Local Similarity 75.0%; Pred. No. 3.3e+03;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RXHPOPER 8
 DB 899 RDHPOVER 906
 RESULT 14
 AAMS9212
 ID AAMS9212 standard; peptide; 8 AA.
 XX
 AC AAMS9212;
 XX
 DT 27-AUG-1998 (first entry)
 XX
 DE Streptavidin tagged peptide ligand #2.
 XX
 KW Streptavidin; ligand; binding affinity; mutant; isolation; purification;
 KW recover; immobilise.
 XX
 OS Synthetic.
 XX
 PN EP835934-A2.
 XX
 PD 15-APR-1998.
 XX
 PF 09-OCT-1997; 97EP-00117504.
 XX
 PR 10-OCT-1996; 96DE-01041876.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH.
 XX
 PI Skerra A, Voss S;
 XX
 DR WPI; 1998-218868/20.
 XX
 PT Streptavidin mutants with higher binding affinity for peptide ligands -

PT have mutation in amino acid region 44-53, used to isolate, purify or
 PT determine fusion proteins including these ligands.

PS Claim 10, Page 11, 21pp; German.

CC AAW59211 and AAW59212 are ligands used in a method to assay binding
 CC affinity of streptavidin mutants. These mutants have a mutation within
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
 CC binding affinity than the wild-type for peptide ligands that include the
 CC sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z
 CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
 CC mutants can be used to isolate, purify and determine proteins or to
 CC determine/recover substances that contain streptavidin-binding groups.
 CC Such compounds may also be used to immobilise fusions on microtitre
 CC plates, microbeads or sensor chips

XX SQ Sequence 8 AA;

Query Match 81.0%; Score 34; DB 2; Length 8;
 Best Local Similarity 71.4%; Pred. No. 1.8e+06;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8
 :|||||:
 2 SHPQPEK 8

RESULT 15

AB35433 ID AAB35433 standard; peptide; 8 AA.

XX AAB35433;

DT 23-MAY-2001 (first entry)

XX Epitope peptide #3.

KM Nascent protein detection; protein analysis; aminoacylated tRNA;
 KW BODIPY marker; disease diagnosis.

XX Unidentified.

XX WO200114578-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000WO-US023233.

XX 25-AUG-1999; 99US-00382736.

PR 25-AUG-1999; 99US-00382950.

XX (AMBE-) AMBERGEN INC.

XX Pl Rochschild KJ, Gite S, Olejnik J;

DR WPI; 2001-168972/17.

PT Method for detecting nascent proteins by fluorescence comprises
 PT misaminoacylating a tRNA molecule with a marker compound, useful for
 PT detecting mutations in proteins, e.g. cancer.

PS Disclosure; Page 47; 204pp; English.

CC The present invention describes a method of detecting nascent proteins
 CC involving aminoacylating a tRNA molecule with a 4',4'-difluoro-4-bora-3a,4a
 CC -diaz-a-8-indacene (BODIPY) marker leading to the production of a
 CC misaminoacylated tRNA. This enables the detection, isolation and analysis
 CC of nascent proteins using UV without the usual accompanying radioactivity
 CC problems. It may be used to detect mutations, for example in cancer,
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer

XX Sequence 8 AA;

Query Match 81.0%; Score 34; DB 4; Length 8;
 Best Local Similarity 71.4%; Pred. No. 1.8e+06;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8
 :|||||:
 DB 2 SHPQPEK 8

Search completed: March 2, 2005, 13:02:46
 Job time : 36.0325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005: CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 / Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ9
Perfect score: 42
Sequence: 1 rxhpofcr 8

Scoring table: BLOSUM62DX
Gapop 10.0, Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	231	15	US-10-104-047-3576
2	39	92.9	359	15	US-10-369-493-20354
3	39	92.9	360	15	US-10-369-493-1288
4	37	88.1	362	15	US-10-369-493-21610
5	36	85.7	619	15	US-10-282-122A-47374
6	35	83.3	36	14	US-10-026-578B-11
7	35	83.3	2747	15	US-10-402-842-2
8	34	81.0	8	9	US-09-809-517A-9
9	34	81.0	8	9	US-09-973-145-7
10	34	81.0	8	12	US-09-813-197-8
11	34	81.0	8	13	US-10-104-218-5
12	34	81.0	8	13	US-10-208-357-9
13	34	81.0	8	14	US-10-001-934-8

14	34	81.0	8	14	US-10-026-578B-2	Sequence 2, Appli
15	34	81.0	8	14	US-10-026-578B-9	Sequence 9, Appli
16	34	81.0	8	14	US-10-174-368A-7	Sequence 7, Appli
17	34	81.0	8	14	US-10-264-127-8	Sequence 8, Appli
18	34	81.0	8	15	US-10-339-712-8	Sequence 8, Appli
19	34	81.0	8	15	US-10-339-712-8	Sequence 67, Appli
20	34	81.0	8	15	US-10-275-046-4	Sequence 4, Appli
21	34	81.0	8	15	US-10-425-000-76	Sequence 76, Appli
22	34	81.0	8	15	US-10-424-999-25	Sequence 25, Appli
23	34	81.0	8	15	US-10-358-283-23	Sequence 23, Appli
24	34	81.0	8	16	US-10-628-432-11	Sequence 41, Appli
25	34	81.0	8	17	US-10-494-248-17	Sequence 17, Appli
26	34	81.0	8	17	US-10-634-645-11	Sequence 11, Appli
27	34	81.0	8	17	US-10-719-523-8	Sequence 8, Appli
28	34	81.0	9	9	US-09-983-067-3	Sequence 3, Appli
29	34	81.0	10	9	US-09-809-517A-6	Sequence 6, Appli
30	34	81.0	10	15	US-10-147-211A-20	Sequence 20, Appli
31	34	81.0	11	15	US-10-354-983-29	Sequence 29, Appli
32	34	81.0	11	16	US-10-628-432-25	Sequence 25, Appli
33	34	81.0	13	15	US-10-338-592-30	Sequence 30, Appli
34	34	81.0	21	9	US-09-809-517A-30	Sequence 30, Appli
35	34	81.0	22	9	US-09-809-517A-31	Sequence 31, Appli
36	34	81.0	22	9	US-09-809-517A-31	Sequence 31, Appli
37	34	81.0	24	14	US-10-026-578B-4	Sequence 3, Appli
38	34	81.0	24	14	US-10-026-578B-4	Sequence 4, Appli
39	34	81.0	25	9	US-09-809-517A-34	Sequence 34, Appli
40	34	81.0	36	14	US-10-026-578B-10	Sequence 10, Appli
41	34	81.0	117	10	US-09-977-137A-4	Sequence 4, Appli
42	34	81.0	117	10	US-09-977-137A-5	Sequence 5, Appli
43	34	81.0	117	10	US-09-977-137A-7	Sequence 7, Appli
44	34	81.0	117	10	US-09-977-137A-8	Sequence 8, Appli
45	34	81.0	117	10	US-09-977-137A-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-10-104-047-3576
; Sequence 3576, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3576
; TYPE: PRT
; LENGTH: 231
; ORGANISM: Homo sapiens
US-10-104-047-3576

Query Match 92.9%; Score 39; DB 15; Length 231;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 RXHPQFER 8
DB 100 RXHPQFER 107
US-10-369-493-20354
; Sequence 20354, Application US/10369493
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

```

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20354
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20354

Query Match          92.9%; Score 39; DB 15; Length 359;
Best Local Similarity 75.0%; Pred. No. 77;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
       |||:|
Db      311 RKHPFER 318

RESULT 3
US-10-369-493-1288
; Sequence 1288, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374.
; SEQ ID NO 1288
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1288

Query Match          92.9%; Score 39; DB 15; Length 360;
Best Local Similarity 75.0%; Pred. No. 77;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
       |||:|
Db      311 RKHPFER 318

RESULT 4
US-10-369-493-21610
; Sequence 21610, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493

```

```

; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21610
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21610

Query Match          88.1%; Score 37; DB 15; Length 362;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
       |||:|
Db      311 RKHPFER 318

RESULT 5
US-10-282-122A-47374
; Sequence 47374, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47374
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47374

Query Match          85.7%; Score 36; DB 15; Length 619;
Best Local Similarity 75.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```


QY 1 RXHPOFER 8
|:|||||
Db 339 RHPOLTER 346.

RESULT 6
US-10-026-578B-11
; Sequence 11, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tags
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: MISC FEATURE
; LOCATION: (9)_(128)
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11

Query Match 83.3%; Score 35; DB 14; Length 36;
Best Local Similarity 62.5%; Pred. No. 41;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 8
|:|||||
Db 29 QSHPOFEK 36

RESULT 7
US-10-402-842-2
; Sequence 2, Application US/10402842
; Publication No. US20030219872A1
; GENERAL INFORMATION:
; APPLICANT: Magarey, Nathan A.
; APPLICANT: Hucul, John A.
; TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GENES
; FILE REFERENCE: 0630/17854-US1
; CURRENT APPLICATION NUMBER: US/10/402,842
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US 60/368,713
; PRIOR FILING DATE: 2002-03-30
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2747
; TYPE: PRT
; ORGANISM: Streptomyces hygroscopicus
US-10-402-842-2

Query Match 83.3%; Score 35; DB 15; Length 2747;
Best Local Similarity 75.0%; Pred. No. 2,8e+03;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8
|:|||||
Db 899 RHPOLTER 906

RESULT 8
US-09-809-517A-9
; Sequence 9, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/proteins
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match 81.0%; Score 34; DB 9; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
|:|||||
Db 2 SHPOFEK 8

RESULT 9
US-09-973-145-7
; Sequence 7, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sedana
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-06819
; CURRENT APPLICATION NUMBER: US/09/973,145
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 09/382,950
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-973-145-7

Query Match 81.0%; Score 34; DB 9; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
|:|||||
Db 2 SHPOFEK 8

```
RESULT 10
US-09-813-197-8
; Sequence 8, Application US/09813197
; Publication No. US2005009013A1
; GENERAL INFORMATION:
; APPLICANT: Rothenchild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBER-03951
; CURRENT APPLICATION NUMBER: US/09/813,197
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-813-197-8
```

```
Query Match      81.0%; Score 34; DB 12; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPOFER 8
       :|||||:
Db      2 SHPOFEK 8
```

```
RESULT 11
US-10-104-218-5
; Sequence 5, Application US/10104218
; Publication No. US20020177196A1
; GENERAL INFORMATION:
; APPLICANT: MAIER, Thomas
; APPLICANT: GABBERT, Carsten
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
; FILE REFERENCE: MAIER, T. ET AL.-2
; CURRENT APPLICATION NUMBER: US/10/104,218
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match      81.0%; Score 34; DB 13; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPOFER 8
       :|||||:
Db      2 SHPOFEK 8
```

```
RESULT 12
US-10-208-357-9
; Sequence 9, Application US/10208357
; Publication No. US20020182687A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Kurz, Markus
; APPLICANT: Lohse, Peter
; APPLICANT: Wagner, Richard
; TITLE OF INVENTION: Peptide Acceptor Ligation Methods
; FILE REFERENCE: 50036/031002
; CURRENT APPLICATION NUMBER: US/10/208,357
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: US/09/619,103
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,834
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9
```

```
Query Match      81.0%; Score 34; DB 13; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPOFER 8
       :|||||:
Db      2 SHPOFEK 8
```

```
RESULT 13
US-10-001-934-8
; Sequence 8, Application US/10001934
; Publication No. US20030032782A1
; GENERAL INFORMATION:
; APPLICANT: NAGY, ZOLTAN
; TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
; FILE REFERENCE: GRCG-P01-003
; CURRENT APPLICATION NUMBER: US/10/001,934
; CURRENT FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8
```

```
Query Match      81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPOFER 8
       :|||||:
Db      2 SHPOFEK 8
```

```
RESULT 14
US-10-026-578B-2
; Sequence 2, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.010U1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
```

```

; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

```

```

Query Match      81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 XHPQFER 8
:|||||:
Db      2 SHPOFEK 8

```

```

RESULT 15
US-10-026-578B-9
; Sequence 9, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

```

```

Query Match      81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 XHPQFER 8
:|||||:
Db      2 SHPOFEK 8

```

```

Search completed: March 2, 2005, 14:18:51
Job time : 24.878 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds

(without alignments)
65.585 Million cell updates/sec

Title: SEQ9

Perfect score: 42

Sequence: 1 rxhpgfer 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/6CTUS.COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	85.7	482	4	US-09-107-532A-4868
2	35	83.3	128	4	US-09-270-767-32326
3	35	83.3	128	4	US-09-270-767-47543
4	34	81.0	8	3	US-08-948-097-2
5	34	81.0	8	3	US-09-382-950-7
6	34	81.0	8	3	US-09-382-736B-8
7	34	81.0	8	4	US-09-619-103-9
8	34	81.0	8	4	US-10-104-218-5
9	34	81.0	8	4	US-09-809-517A-9
10	34	81.0	10	4	US-09-809-517A-6
11	34	81.0	13	2	US-08-687-559-30
12	34	81.0	13	4	US-09-401-415-30
13	34	81.0	21	4	US-09-809-517A-30
14	34	81.0	22	4	US-09-809-517A-33
15	34	81.0	24	4	US-09-809-517A-31
16	34	81.0	25	4	US-09-809-517A-34
17	34	81.0	117	4	US-09-977-137A-4
18	34	81.0	117	4	US-09-977-137A-5
19	34	81.0	117	4	US-09-977-137A-7
20	34	81.0	117	4	US-09-977-137A-8
21	34	81.0	117	4	US-09-977-137A-9
22	34	81.0	117	4	US-09-977-137A-10
23	34	81.0	117	4	US-09-977-137A-11
24	34	81.0	117	4	US-09-977-137A-12
25	34	81.0	118	4	US-09-977-137A-6
26	34	81.0	275	4	US-09-270-767-58917
27	34	81.0	777	4	US-09-902-540-9844

28	34	81.0	801	4	US-09-270-767-43549	Sequence 43549, A
29	34	81.0	3666	2	US-08-222-617A-12	Sequence 12, Appl
30	34	81.0	3727	2	US-08-222-617A-27	Sequence 27, Appl
31	34	81.0	3778	2	US-08-222-617A-2	Sequence 2, Appl
32	33	78.6	103	4	US-09-270-767-41422	Sequence 41422, A
33	33	78.6	114	4	US-09-270-767-42719	Sequence 42719, A
34	33	78.6	205	4	US-09-252-991A-25397	Sequence 25397, A
35	33	78.6	369	4	US-09-252-991A-17585	Sequence 17585, A
36	33	78.6	413	4	US-09-723-546-11	Sequence 11, Appl
37	33	78.6	444	1	US-09-178-002-2	Sequence 17, Appl
38	33	78.6	466	3	US-08-704-711A-17	Sequence 17, Appl
39	33	78.6	466	3	US-09-521-220-17	Sequence 17, Appl
40	33	78.6	467	1	US-09-178-002-4	Sequence 4, Appl
41	33	78.6	467	3	US-09-391-104-24	Sequence 24, Appl
42	33	78.6	468	3	US-08-448-489-13	Sequence 13, Appl
43	33	78.6	468	3	US-09-689-730-13	Sequence 13, Appl
44	33	78.6	477	4	US-09-252-991A-16778	Sequence 16778, A
45	33	78.6	650	4	US-09-107-532A-5521	Sequence 5521, Ap

ALIGNMENTS

RESULT 1
US-09-107-532A-4868
Sequence 4868, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Malham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4868:
SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...482
SEQUENCE DESCRIPTION: SEQ ID NO: 4868:
US-09-107-532A-4868

```

Query Match      85.7%; Score 36; DB 4; Length 482;
Best Local Similarity 62.5%; Pred. No. 45;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
       :|:|:|:|
DB      409 KKHPEFER 416

RESULT 2
US-09-270-767-32326
; Sequence 32326, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32326
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-32326

Query Match      83.3%; Score 35; DB 4; Length 128;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
       :|:|:|:|
DB      43 RSHPOFER 50

RESULT 3
US-09-270-767-47543
; Sequence 47543, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47543
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-47543

Query Match      83.3%; Score 35; DB 4; Length 128;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
       :|:|:|:|
DB      43 RSHPOFER 50

RESULT 4
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skeira, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Muteins
; FILE REFERENCE: HUBR 1119

```

```

; CURRENT APPLICATION NUMBER: US/08/948,097C
; CURRENT FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match      81.0%; Score 34; DB 3; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOFER 8
       :|:|:|:|
DB      2 SHPOFER 8

RESULT 5
US-09-382-950-7
; Sequence 7, Application US/09382950
; Patent No. 6303337
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerry
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMER-03879
; CURRENT APPLICATION NUMBER: US/09/382,950
; CURRENT FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial/Unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: ( )
; OTHER INFORMATION: Synthetic
US-09-382-950-7

Query Match      81.0%; Score 34; DB 3; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOFER 8
       :|:|:|:|
DB      2 SHPOFER 8

RESULT 6
US-09-382-736B-8
; Sequence 8, Application US/09382736B
; Patent No. 6306628
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerry
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMER-03951
; CURRENT APPLICATION NUMBER: US/09/382,736B
; CURRENT FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT

```

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-382-73B-8

Query Match 81.0%; Score 34; DB 3; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHOPER 8
Db 2 SHPOEX 8

RESULT 7
US-09-619-103-9
Sequence 9, Application US/09619103
Patent No. 6429300
GENERAL INFORMATION:
APPLICANT: Kurz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Wagner, Richard
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/09/619,103
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed sequence to act as an identifying tag
US-09-619-103-9

Query Match 81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHOPER 8
Db 2 SHPOEX 8

RESULT 8
US-10-104-218-5
Sequence 5, Application US/10104218
Patent No. 6579705
GENERAL INFORMATION:
APPLICANT: Maier, Thomas
APPLICANT: Gaeber, Carsten
TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
FILE REFERENCE: MAIER, T. ET AL.-2
CURRENT APPLICATION NUMBER: US/10/104,218
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
PRIOR FILING DATE: 2001-05-03
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(8)
OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5

Query Match 81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHOPER 8
Db 2 SHPOEX 8

RESULT 9
US-09-809-517A-9
Sequence 9, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match 81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHOPER 8
Db 2 SHPOEX 8

RESULT 10
US-09-809-517A-6
Sequence 6, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 10
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-6

Query Match 81.0%; Score 34; DB 4; Length 10;
Best Local Similarity 71.4%; Pred. No. 1.5;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8
:|||||:
Db 4 SHPOFER 10

RESULT 11

US-08-687-559-30
Sequence 30, Application US/08687559
Patent No. 5955647
GENERAL INFORMATION:
APPLICANT: Fitchen, John H.
APPLICANT: Beachy, Roger N.
TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
TITLE OF INVENTION: OVERPRODUCE PEPTIDES AND PROTEINS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,559
FILING DATE: NO. 5955647ember 18, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01467
FILING DATE: 03-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: learn, June M.
REGISTRATION NUMBER: 31,238
REFERENCE/DOCKET NUMBER: 07302/011001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-687-559-30

Query Match 81.0%; Score 34; DB 2; Length 13;
Best Local Similarity 71.4%; Pred. No. 2;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8
:|||||:
Db 7 GHPOFOR 13

RESULT 12
US-09-401-415-30
Sequence 30, Application US/09401415
Patent No. 6503732
GENERAL INFORMATION:
APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
OVERPRODUCE PEPTIDES AND PROTEINS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1860 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA

ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/401,415
FILING DATE: 21-Sep-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01467
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Bostich, June M.
REGISTRATION NUMBER: 31,238
REFERENCE/DOCKET NUMBER: PD-4074
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..13
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-401-415-30

Query Match 81.0%; Score 34; DB 4; Length 13;
Best Local Similarity 71.4%; Pred. No. 2;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8
:|||||:
Db 7 GHPOFOR 13

RESULT 13
US-09-809-517A-30
Sequence 30, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 30
LENGTH: 21
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match 81.0%; Score 34; DB 4; Length 21;
Best Local Similarity 71.4%; Pred. No. 3.4;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8
:|||||:

Db 15 SHPOFEK 21

RESULT 14

```

US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

```

Query Match

Best Local Similarity 81.0%; Score 34; DB 4; Length 22;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8

Db 16 SHPOFEK 22

RESULT 15

```

US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

```

Query Match

Best Local Similarity 81.0%; Score 34; DB 4; Length 24;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8

Db 18 SHPOFEK 24

Search completed: March 2, 2005, 12:25:35
Job time : 9.10569 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 15.2846 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ10
Perfect score: 124
Sequence: 1 wshpgefkgsgswshpgefek 20

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	46.8	493	2	B72704
2	52	41.9	103	2	C85713
3	52	41.9	103	2	A90904
4	50	40.3	346	1	S35500
5	49	39.5	726	2	T35257
6	47	37.9	103	2	E90973
7	47	37.9	103	2	A85821
8	47	37.9	433	2	H83444
9	47	37.9	472	2	AH2925
10	47	37.9	472	2	F98356
11	46.5	37.5	1589	1	RGHYC5
12	46	37.1	233	1	C48560
13	46	37.1	243	3	F90197
14	46	37.1	385	2	D41732
15	46	37.1	1885	2	JC2183
16	45.5	36.7	256	2	JU0268
17	45.5	36.7	256	2	S37149
18	45.5	36.7	256	2	A54281
19	45	36.3	391	2	A44063
20	45	36.3	418	2	T47818
21	45	36.3	517	2	S14205
22	45	36.3	580	2	H86189
23	45	36.3	718	2	C90555
24	45	36.3	725	2	S60712
25	44.5	35.9	161	2	B42827
26	44.5	35.9	499	2	JC6141
27	44.5	35.9	759	2	D81657
28	44.5	35.9	83	2	T17752
29	44	35.5	83	2	T17752

30	44	35.5	330	2	AC2071	hypothetical prote
31	44	35.5	342	2	S37596	protein kinase MCP
32	44	35.5	365	2	AD1634	X-Pro dipeptidase
33	44	35.5	507	1	QDBE41	BGLF1 protein - hu
34	44	35.5	1253	3	T21065	hypothetical prote
35	43.5	35.1	597	1	HQEC1	hydrogenase (BC 1.
36	43.5	35.1	597	2	E85632	hydrogenase-1 larg
37	43.5	35.1	597	2	A90770	hydrogenase-1 larg
38	43	34.7	55	2	I78887	hypothetical NF-1
39	43	34.7	68	2	E85918	hypothetical prote
40	43	34.7	172	2	D82642	conserved hypothet
41	43	34.7	187	2	A45878	hypothetical prote
42	43	34.7	209	2	C89005	protein T24A6.3 (I
43	43	34.7	216	2	G75128	hypothetical prote
44	43	34.7	234	2	D96592	hypothetical prote
45	43	34.7	251	2	C81816	hypothetical prote

ALIGNMENTS

RESULT 1
B72704
hypothetical protein APE1049 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: B72704
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Maeda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: B72704
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-493 <KAW>
A:Cross-references: UNIPROT:Q9YD63; DDBJ:AP000060; NID:G5104188; PIDN:BAAB0034.1; PID:dl A:Experimental source: strain K1
A:Gene: APE1049

Query Match
Best Local Similarity 46.8%; Score 58; DB 2; Length 493;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 WSHPGEFKGSGWSHPGEF 19
DB 333 WPNPLFVVGGTWEHVIE 351

RESULT 2
C85713
unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1 C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: C85713
R:Perera, N.T.; Plunkett II, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimmlante, E.; Potlamousis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: C85713
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:O8X549; GB:AE005174; NID:G12515046; PIDN:AGS6167.1; GSPDB: C:Experimental source: strain O157:H7, substrain EDL933
A:Gene: Z2097

Query Match
Best Local Similarity 41.9%; Score 52; DB 2; Length 103;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 POFKGGGSMHPQEK 20
DB 8 PVRDGYGYWTHPEYK 24

RESULT 3
A:Accession: A90904
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A90904
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAY>
A:Cross-references: UNIPROT:Q8X549; GB:BA000007; PIDN:BA835624.1; PID:G13361667; GSPDB:G
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs2201

Query Match 41.9%; Score 52; DB 2; Length 103;
Best Local Similarity 47.1%; Pred. No. 1.4;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 4 POFKGGGSMHPQEK 20
DB 8 PVRDGYGYWTHPEYK 24

RESULT 4
A:Accession: S35500
C:Species: Caenorhabditis elegans
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
R:Itawaaki, M.; Okumura, K.; Kondo, Y.; Tanaka, T.; Igarashi, H.
Nucleic Acids Res. 20, 4001-4007, 1992
A>Title: CDNA cloning of a novel heterogeneous nuclear ribonucleoprotein gene homologue
A:Reference number: S35500; MUID:923755684; PMID:1354852
A:Accession: S35500
A:Molecule type: mRNA
A:Residues: 1-346 <IMA>
A:Cross-references: UNIPROT:Q22037; EMBL:S43152
R:Du, Z.; Scheet, P.; Andrews, S.
Submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F42A6.
A:Reference number: Z21201
A:Accession: T32620
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-346 <DUZ>
A:Cross-references: EMBL:AF038613; PIDN:AA892051.1; GSPDB:GN00022; CESP:F42A6.7
A:Experimental source: strain Bristol N2; clone F42A6
C:Genetics:
A:Gene: CESP:F42A6.7
A:Map position: 4
A:Introns: 9/3; 255/1
C:Superfamily: helix-stabilizing protein; ribonucleoprotein repeat homology
F:24-90/Domain: ribonucleoprotein repeat homology <RRM1>
F:115-161/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 40.3%; Score 50; DB 1; Length 346;
Best Local Similarity 53.8%; Pred. No. 10;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 WSHPOFKGGGSM 13

DB 271 WGGPQQQGGGGM 283

RESULT 5
A:Accession: T35257
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
R:Oliver, K.; Harris, D.; Bentley, S.D.; Partridge, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T35257
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-726 <OLI>
A:Cross-references: UNIPROT:Q9X7N8; EMBL:AL049587; PIDN:CAB40682.1; GSPDB:GN00070; SCOE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOE:SCSFP2.15
C:Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de
F:321-605/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>

Query Match 39.5%; Score 49; DB 2; Length 726;
Best Local Similarity 44.4%; Pred. No. 31;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY 1 WSHPOFKGGGSMHPQEK 18
DB 217 WDRPGYRIPGTPSPHPR 234

RESULT 6
A:Accession: E90973
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAY>
A:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BA836180.1; PID:G13362225; GSPDB:G
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs2757

Query Match 37.9%; Score 47; DB 2; Length 103;
Best Local Similarity 60.0%; Pred. No. 7.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 11 GWSHPOFK 20
DB 15 GCVTHPEYK 24

RESULT 7
A:Accession: A85821
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimataanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A95821
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <STO>
 A:Crosss-references: UNIPROT:Q8X4V0; GB:AE005174; NID:g12516136; PIDN:AA657029.1; GSPDB:G
 A:Experimental source: strain 0157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z3120

Query Match 37.9%; Score 47; DB 2; Length 103;
 Best Local Similarity 60.0%; Pred. No. 7.3;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Oy 11 GMSHPOFEK 20
 Db 15 GCMTHPEYK 24

RESULT 8
 H83444
 Hypothetical protein c PA1600 (imported) - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: H83444
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: H83444
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-433 <STO>
 A:Crosss-references: UNIPROT:Q913C1; GB:AE004588; GB:AE004091; NID:g9947563; PIDN:AA60498
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA1600

C:Superfamily: membrane-bound alcohol dehydrogenase, cytochrome c subunit; cytochrome c
 C:Keywords: chromoprotein; heme; iron; metalloprotein
 F:58, 61/Binding site: heme (Cys) (covalent) #status predicted
 F:62/Binding site: heme iron (His) (axial ligand) #status predicted
 F:204, 207/Binding site: heme (Cys) (covalent) #status predicted
 F:208/Binding site: heme iron (His) (axial ligand) #status predicted
 F:336, 339/Binding site: heme (Cys) (covalent) #status predicted
 F:340/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 37.9%; Score 47; DB 2; Length 433;
 Best Local Similarity 47.1%; Pred. No. 34;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 Oy 4 POFKGGSMHPOFEK 20
 Db 90 PDRAAGIGMSYPAPER 106

RESULT 9
 AH2925
 Hypothetical protein Atu3006 (imported) - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
 C:Accession: AH2925
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
 erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell
 i Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Teo, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 eier, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AH2925
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-472 <KIR>
 A:Crosss-references: UNIPROT:Q8UBU0; GB:AE008669; PIDN:AA43822.1; PID:g17741363; GSPDB:G
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu3006
 A:Map position: linear chromosome

Query Match 37.9%; Score 47; DB 2; Length 472;
 Best Local Similarity 63.6%; Pred. No. 37;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Oy 5 OFKGGSMH 15
 Db 202 RPKAGGLMTH 212

RESULT 10
 P98356
 Hypothetical protein AGR_L_3586 (imported) - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: P98356
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.,
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: P98356
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-472 <KIR>
 A:Crosss-references: UNIPROT:Q8UBU0; GB:AE007870; PIDN:AAK90376.1; PID:g15160420; GSPDB:G
 C:Genetics:
 A:Gene: AGR_L_3586
 A:Map position: linear chromosome

Query Match 37.9%; Score 47; DB 2; Length 472;
 Best Local Similarity 63.6%; Pred. No. 37;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Oy 5 OFKGGSMH 15
 Db 202 RPKAGGLMTH 212

RESULT 11
 R6BYC5
 cell division control protein CDC25 - Yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein L2142.6; protein YJR310c
 C:Species: Saccharomyces cerevisiae
 C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
 C:Accession: A26596; S51442; A23444; S43051; S47990
 R:Broek, D.; Toda, T.; Michaeli, T.; Levin, L.; Birnmeier, C.; Zoller, M.; Powers, S.;
 Cell 48, 789-799, 1987
 A:Title: The S. cerevisiae CDC25 gene product regulates the RAS/adenylylate cyclase pathwa
 A:Reference number: A26596; MUID:87131091; PMID:3545497
 A:Accession: A26596
 A:Molecule type: DNA
 A:Residues: 1-1589 <BRO>
 A:Crosss-references: UNIPROT:P04821; EMBL:M15458; NID:g171184; PIDN:AAA34478.1; PID:g1711
 R:Pauley, A.
 submitted to the EMBL Data Library, November 1994
 A:Description: The sequence of S. cerevisiae comid L2142.
 A:Reference number: S51437
 A:Accession: S51442
 A:Molecule type: DNA
 A:Residues: 1-1589 <PAU>
 A:Crosss-references: EMBL:U17247; NID:g577216; PIDN:AA67360.1; PID:g577222; GSPDB:GN0001
 R:Camonis, J.H.; Kalekine, M.; Gondre, B.; Garreau, H.; Boy-Marcotte, E.; Jacquet, M.
 EMBO J. 5, 375-380, 1986
 A:Title: Characterization, cloning and sequence analysis of the CDC25 gene which control
 A:Reference number: A23444; MUID:86220116; PMID:3011405
 A:Accession: A23444

A:Molecule type: DNA
 A:Residues: 1-496, Y, 498-953, 'LSVIMLSR', 964-1589 <CAM>
 A:Cross-references: EMBL:X03579; NID:93483; PIDN:CAA27259.1; PID:93484
 R:Daniel, J.H.
 Curr. Genet. 10, 879-885, 1986
 A:Title: The CDC25 "Start" gene of *Saccharomyces cerevisiae*: sequencing of the active C-terminus
 A:Reference number: S43051; MUID:88194639; PMID:3329037
 A:Accession: S43051
 A:Molecule type: DNA
 A:Residues: 877-1589 <DAN>
 A:Cross-references: EMBL:X03579
 C:Genetics:
 A:Gene: SGD: CDC25; CTN1; MIPS: YLR310C
 A:Cross-references: SGD: S0004301; MIPS: YLR310C
 A:Map position: 12R
 C:Function:
 A:Description: positive control of level of cellular cAMP at the stage at which the cell
 C:Superfamily: budding yeast CDC25; CDC25-type guanine nucleotide exchange activator
 C:Keywords: cell cycle control; transmembrane protein
 P:55-123/Domain: BH3 homology <SH3>
 P:1301-1542/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 37.1%; Score 46.5; DB 1; Length 1589;
 Best Local Similarity 56.2%; Pred. No. 1.6e+02;
 Matches 9; Conservative 3; Mismatches 1; Indels 3; Gaps 1;

QY 4 HPQFKGSGSWSHP 16
 DB 699 HPQFKGSGSWSHP 714

RESULT 12
 C48560
 UL56 protein - human herpesvirus 1 (strain HFEM)
 C:Species: human herpesvirus 1
 C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jul-2004
 C:Accession: C48560
 R:Rosen-Wolff, A.; Frank, S.; Raab, K.; Moyal, M.; Becker, Y.; Darai, G.
 Virus Res. 25, 189-199, 1992
 A:Title: Determination of the coding capacity of the BamHI DNA fragment B of apathogenic
 A:Reference number: A48560; MUID:93070559; PMID:1332274
 A:Accession: C48560
 A:Molecule type: DNA
 A:Residues: 1-233 <ROS>
 A:Cross-references: UNIPROT:P36297
 A:Note: sequence extracted from NCBI backbone (NCBIN:117573, NCBI:P117577)
 C:Genetics:
 A:Gene: UL56
 C:Superfamily: herpesvirus UL56 protein

Query Match 37.1%; Score 46; DB 1; Length 233;
 Best Local Similarity 43.8%; Pred. No. 24;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 WSHPOFKGSGSWSHP 16
 DB 108 WSHPOFKGSGSWSHP 123

RESULT 13
 P90197
 hypothetical protein SSO0519 [imported] - *Sulfolobus solfataricus*
 C:Species: *Sulfolobus solfataricus*
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
 C:Accession: P90197
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awaizer, M.J.; Chan, J.; Jongs, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.; Jongs, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: *Sulfolobus solfataricus* complete genome.
 A:Reference number: A99139
 A:Accession: P90197
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-243 <KUR>
 A:Cross-references: UNIPROT:Q9UWS5; GB:AE006641; NID:913813677; PIDN:AAK40837.1; GSPDB:C
 C:Genetics:
 A:Gene: SSO0519

Query Match 37.1%; Score 46; DB 2; Length 243;
 Best Local Similarity 77.8%; Pred. No. 26;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 HPQFKGSGG 11
 DB 28 HPQFKGSGG 36

RESULT 14
 D41732
 heterogeneous nuclear RNP protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C>Date: 04-Mar-1993 #sequence_revision 02-Aug-1994 #text_change 16-Aug-2004
 C:Accession: D41732; G48110
 R:Matunis, E.L.; Matunis, M.J.; Dreyfuss, G.
 J. Cell Biol. 116, 257-269, 1992
 A:Title: Characterization of the major hnRNP proteins from *Drosophila melanogaster*.
 A:Reference number: A41732; MUID:92112968; PMID:1730754
 A:Accession: D41732
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-385 <MAT>
 A:Cross-references: UNIPROT:P48809; GB:X62639; GB:S76628; NID:911041; PIDN:CAA44505.1; P
 R:Kim, Y.J.; Baker, B.S.
 Mol. Cell. Biol. 13, 174-183, 1993
 A:Title: Isolation of RNA-type RNA-binding protein genes and the analysis of their relat
 A:Reference number: A48110; MUID:93109300; PMID:8417324
 A:Accession: G48110
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 10-53 <KIM>
 A:Cross-references: GB:S51720; NID:9262267; PIDN:AA824628.1; PID:9262268
 A:Note: sequence extracted from NCBI backbone (NCBIN:121150, NCBI:P121151)
 C:Genetics:
 A:Gene: FLYbase:Hb98DE; FLYbase:RBP7
 A:Cross-references: FLYbase:FBgn0001215; FLYbase:FBgn0010261
 C:Superfamily: ribonucleoprotein repeat homology <RBP>
 F:8-74/Domain: ribonucleoprotein repeat homology <RBP>
 F:97-163/Domain: ribonucleoprotein repeat homology <RBP>

Query Match 37.1%; Score 46; DB 2; Length 385;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFKGSGSWS 14
 DB 261 WSHPOFKGSGSWS 274

RESULT 15
 JQ2183
 hypothetical 216.5K protein - apple chlorotic leaf spot virus (isolate P-205)
 N:Alternate names: ORF 1 protein
 C:Species: apple chlorotic leaf spot virus
 C>Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 09-Jul-2004
 C:Accession: JQ2183
 R:Sato, K.; Yoshikawa, N.; Takahashi, T.
 J. Gen. Virol. 74, 1927-1931, 1993
 A:Title: Complete nucleotide sequence of the genome of an apple isolate of apple chlorot
 A:Reference number: JQ2183; MUID:93389448; PMID:8376968
 A:Accession: JQ2183
 A:Molecule type: mRNA
 A:Residues: 1-1885 <SAT>
 A:Cross-references: UNIPROT:P54891; GB:D14996; NID:9434059; PIDN:BA03641.1; PID:9453240
 C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

C:Keywords: ATP

Query Match 37.1%; Score 46; DB 2; length 1885;
Best Local Similarity 43.8%; Pred. No. 2.3e+02;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WSHPOFEKGGGSMSP 16
|||:|:|:|
Db 1434 WSNQPREKGPWKHP 1449

Search completed: March 2, 2005, 12:28:56
Job time : 16.2846 secs

This Page Blank (uspto)


```

RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB097860; BAC44887.1; -.
DR HSP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR001054; G_cyclase.
DR InterPro; IPR009080; tRNAyn_1a_bind.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
DR Lyase.
SQ SEQUENCE 743 AA; 83251 MW; ACFSC53E0982813A CRC64;

QY Query Match 49.2%; Score 61; DB 2; Length 743;
   Best Local Similarity 100.0%; Pred. No. 3.6;
   Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 734 GSWSHPOFEK 743

```

```

RESULT 3
ID Q9YD63 PRELIMINARY; PRT; 493 AA.
AC Q9YD63;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical; protein APEI049.
GN OrderedLocustNames=APEI049;
OS Aeropyrum pernix.
OC Archaea; Cepharchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OC NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kusbida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KL.";
RL DNA Ref. 6:83-101(1999).
DR EMBL; AP000060; BAA80034.1; -.
DR PIR; B72704; B72704.
DR GO; GO:0005381; F:iron ion transporter activity; IEA.
DR GO; GO:0006827; P:high affinity iron ion transport; IEA.
DR InterPro; IPR002491; Peripla_BP.
DR InterPro; IPR009053; Peripla_BP.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR010978; tRNA binding_arm.
DR Pfam; PF01497; Peripla_BP_2; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 493 AA; 52685 MW; 9AA07ED93055826E CRC64;

QY Query Match 46.8%; Score 58; DB 2; Length 493;
   Best Local Similarity 47.4%; Pred. No. 6.1;
   Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 333 WENPLEVVGGTWEHVIE 351

```

RESULT 4
Q8A6Y2

```

ID Q8A6Y2 PRELIMINARY; PRT; 507 AA.
AC Q8A6Y2;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=B71743;
OS Bacteroides thetaiotaomicron.
OC Bacterioides; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OC NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
DR EMBL; AE016933; AA076850.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR01087; Lipase_GDSL.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 507 AA; 56254 MW; B2B890D937A1E78A CRC64;

QY Query Match 45.2%; Score 56; DB 2; Length 507;
   Best Local Similarity 57.1%; Pred. No. 12;
   Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 200 HPEFMPGNDWVHP 213

3 HPOFEKGGSGMSHP 16

RESULT 5
ID Q89NL8 PRELIMINARY; PRT; 1174 AA.
AC Q89NL8;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE B113820 protein.
GN OrderedLocustNames=B113820;
OS Bradyrhizobium japonicum.
OC Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OC NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasaamoto S., Watanabe A., Idegawa K., Iritguchi M., Kawashina K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurutoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Ref. 9:189-197(2002).
DR EMBL; AP005949; BAC49085.1; -.
DR GO; GO:0004022; F:alcohol dehydrogenase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR008274; Aldxan_dh_bind.
DR InterPro; IPR000674; Aldxan_dh_hamm.
DR InterPro; IPR003219; Cytc_ach.
DR InterPro; IPR003345; Cytc_heme_BS.
DR InterPro; IPR009056; Cytochrome_c.
DR InterPro; IPR003088; Cyt_Ct.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF01315; Ald_xan_dh_C; 1.
DR Pfam; PF02738; Ald_xan_dh_C2; 3.

```

```

DR Pfam: PF00034; Cytochrom C; 2.
DR ProDom: PD01584; CytoC adh; 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_3.
DR PROSITE: PS00030; RRM_RNF_1; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 1174 AA; 125102 MW; 1365696EB14A549F CRC64;

Query Match
Best Local Similarity 52.9%; Score 55; DB 2; Length 1174;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 4 POPEKGGSGWSHPQFEK 20
DB 837 PDVERTGIGAMSYPAFER 853

RESULT 6
O7PP48 PRELIMINARY; PRT; 168 AA.
AC O7PP48;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DE ENSANGP00000013599 (Fragment).
GN Name=ENSANG00000011110;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008960; EAA11052.2; -.
FT NON_TER
SQ SEQUENCE 168 AA; 18336 MW; 28F3E857B71210C7 CRC64;

Query Match
Best Local Similarity 43.8%; Score 54; DB 2; Length 168;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 WSHPOFEKGGWSWHP 16
DB 95 VNHPSMNHGSAMNHP 110

RESULT 7
O7WV12 PRELIMINARY; PRT; 955 AA.
AC O7WV12;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Glycine cleavage system P protein.
GN Name=gcvP; OrderedLocustNames=PG1305;
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteriia; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RA DOI=10.1128/JB.185.18.5591-5601.2003;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.L.,
RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T.M., Tallon L., Gray J.,
RA Grainger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA Dewhirst F.E., Fraser C.M.;

```

```

RT "Complete genome sequence of the oral pathogenic bacterium
RT Porphyromonas gingivalis strain W83.";
RT J. Bacteriol. 185:5591-5601(2003).
DR EMBL; AB017176; AAO66378.1; -.
DR TIGR; PG1305; -.
DR GO; GO:0005961; C:glycine dehydrogenase complex (decarboxylat. . .; IEA.
DR GO; GO:0004375; P:glycine dehydrogenase (decarboxylating) act. . .; IEA.
DR GO; GO:0006544; P:glycine metabolism; IEA.
DR InterPro; IPR003437; GDC-P.
DR Pfam; PF02347; GDC-P; 1.
DR TIGRFAMs; TIGR00461; gcvP; 1.
SQ SEQUENCE 955 AA; 105962 MW; BF0E2BBA29183205 CRC64;

Query Match
Best Local Similarity 57.1%; Score 54; DB 2; Length 955;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 HPOFEKGGSGWSHP 16
DB 894 HPOEVTANDWSHP 907

RESULT 8
O72IWO PRELIMINARY; PRT; 447 AA.
AC O72IWO;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Putative rRNA methyl transferase.
GN OrderedLocustNames=TRC1021;
OS Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039).
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermates; Thermaceae;
OC Thermus.
OX NCBI_TaxID=262724;
RN (1)
RP SEQUENCE FROM N.A.
RA Pubmed=15064768;
RA Henne A., Brueggemann H., Raasch C., Wieser A., Hartsech T.,
RA Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R.,
RA Jacobi C., Starkuviene V., Schlenzcek S., Dencker S., Huber R.,
RA Klenk H.-P., Kramer W., Merkl R., Gottschalk G., Fritz H.-J.;
RT "The genome sequence of the extreme thermophile Thermus
RT thermophilus.";
RT Nat. Biotechnol. 22:547-553(2004).
DR EMBL; AE017304; AAS81363.1; -.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR011023; Nop2P.
DR InterPro; IPR006174; Nusb_region.
DR PRINTS; PRO1415; ANKYRIN.
DR ProDom; PD005242; Nusb_region; 1.
DR TIGRFAMs; TIGR00446; nop2P; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 447 AA; 48704 MW; 7C4A84CD337A56EA CRC64;

Query Match
Best Local Similarity 34.5%; Score 53.5; DB 2; Length 447;
Matches 10; Conservative 4; Mismatches 6; Indels 9; Gaps 1;

OY 1 WSH-----POPEKGGSGWSHPQFEK 20
DB 274 WPHRLGEGHFLARFKGGAWSTPRLER 302

RESULT 9
O8X549 PRELIMINARY; PRT; 103 AA.
AC O8X549; Q7ADVO;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)

```

```

DE Hypothetical protein Z2097 (Hypothetical protein Ec2201).
GN OrderedLocusNames=EC2201, z2097;
OS Escherichia coli O157:H7.
OC Bacteriia; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomousis K.,
RA Apodoca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RIMD 050952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22 (2001).
DR EMBL; AE005346; AAG56167.1; -
DR EMBL; AP02557; BAB35624.1; -
DR PIR; A90904; A90904.
DR PIR; C85713; C85713.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11980 MM; 5442AAT91CF29EB6 CRC64;

Query Match
Best Local Similarity 41.9%; Score 52; DB 2; Length 103;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 PPEKGGGWSHPOPEK 20
Db 8 PVRDGYGYWTHPEYEK 24

RESULT 10
Q8PBG8 PRELIMINARY; PRT; 146 AA.
AC Q8PBG8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Molybdopterin-converting factor chain 2.
GN Name=moaB; OrderedLocusNames=XAC1100;
OS Xanthomonas axonopodis (pv. citri).
OC Bacteriia; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farsh C.S., Furian L.R.,
RA Queiroz R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cammaran F., Cardozo J., Chambergro F., Clapina L.P.,
RA Cicerelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos B.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidants J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

```

```

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Teubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AB011739; AAM35974.1; -
DR HSP; P30749; IFN0.
DR GO; GO:0006777; P-Mo-molybdopterin cofactor biosynthesis; IEA.
DR InterPro; IPR003448; Mo_biosynth_MoAE.
DR Pfam; PF02391; MoAE; 1.
KW Complete proteome.
SQ SEQUENCE 146 AA; 16139 MM; 546976F19FA46191 CRC64;

Query Match
Best Local Similarity 41.9%; Score 52; DB 2; Length 146;
Matches 6; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 WSHPOPEKGGWSHPOPE 19
Db 127 WKHEHYLEGDAGWLPBQLQ 145

RESULT 11
Q9LVU2 PRELIMINARY; PRT; 150 AA.
AC Q9LVU2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Cbl|AAP26459.1.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
RT clones."
RL DNA Res. 7:31-63 (2000).
DR EMBL; AB018117; BAA97150.1; -
SQ SEQUENCE 150 AA; 16928 MM; F2FA83C3039E644D CRC64;

Query Match
Best Local Similarity 41.1%; Score 51; DB 2; Length 150;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOPEKGGWSW 13
Db 78 WMEPSYEMGGKW 90

RESULT 12
Q7ID58 PRELIMINARY; PRT; 176 AA.
AC Q7ID58;
DT 05-JUN-2004 (TReMBLrel. 27, Created)
DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE CG13011 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=22867302; PubMed=14525923; DOI=10.1101/gr.1311003;
RA Domazet-Losed T., Tautz R.;
RT "An evolutionary analysis of orphan genes in Drosophila.";
RL Genome Res. 13:2213-2219(2003).
DR EMBL; AF532012; AAQ09911.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008253; Marvel.
DR Pfam; PF01284; MARVEL; 1.
FT NON_TER 1 1
SQ SEQUENCE 176 AA; 19490 MW; 33E99FA9F347EB99 CRC64;

Query Match 41.1%; Score 51; DB 2; Length 176;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGGWSHP 16
Db 130 WLHPQFLVRPGFWAP 145

RESULT 13
Q9VXD1 PRELIMINARY; PRT; 176 AA.
AC Q9VXD1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG13011-PA.
GN ORFNames=CG13011;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Chame Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodeon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foaier C., Gabrilian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Miliushina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazonio M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

```

RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker J.S., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Fries E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleib J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Fries E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Belencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS003503; AAF4643.1; -
DR FlyBase; FBgn0030771; CG13011.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008253; Marvel.
DR Pfam; PF01284; MARVEL; 1.
SQ SEQUENCE 176 AA; 19460 MW; FDCAL1AAD316A1D1 CRC64;

Query Match 41.1%; Score 51; DB 2; Length 176;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGGWSHP 16
Db 130 WLHPQFLVRPGFWAP 145

RESULT 14
Q654S7 PRELIMINARY; PRT; 55 AA.
AC Q654S7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein B1135C02.12 (Hypothetical protein
DE OSUBA0011P19.40).
GN Name=B1135C02.12; Synonyms=OSUBA0011P19.40;
OC Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

```

OC  Eubacteriaceae; Oryzae; Oryza.
OX  NCBI_TaxID=39947;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  PubMed=12447438;
RA  Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
   Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
   Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
   Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
   Hishikawa S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
   Ikono M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
   Karsawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
   Machita K., Maehara T., Mizuno H., Miyabayashi T., Mukai Y.,
   Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
   Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
   Shiohara T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
   Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshinara R., Yukawa K.,
   Zhao H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
   Yano M., Jiang J., Gojobori T.;
RA  "The genome sequence and structure of rice chromosome 1.";
RT  Nature 420:312-316(2002).
RL  Nature 420:312-316(2002).
DR  EMBL; AP04358; BAD45690.1; -
   EMBL; AP003211; BAD45055.1; -
   KW  Hypothetical protein.
SQ  SEQUENCE 55 AA; 5354 MW; FBE1DB6BEB6D433C CRC64;

Query Match      40.3%; Score 50; DB 2; Length 55;
Best Local Similarity 50.0%; Pred. No. 7.9;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY  1 WSHPOFEKGGSGWSHP 16
   | : | | | | | | | |
DB  28 WAAGRGEGGGGSAAP 43

RESULT 15
O8PBW7
ID  O8PBW7; PRELIMINARY; PRT; 145 AA.
DT  01-OCT-2002 (TrEMBLrel. 22, Created)
DT  01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  MolYbdopterin-converting factor chain 2.
GN  Name=moae; OrderedLocustNames=XCC0999;
OS  Xanthomonas campestris (pv. campestris).
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
   Xanthomonadaceae; Xanthomonas.
OX  NCBI_TaxID=340;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=ATCC 33913 / NCPPB 528;
   MEDLINE=2022145; PubMed=12024217; DOI=10.1038/417459a;
   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
   Quesglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
   Camarotte G., Canavan F., Cardozo J., Chamberg F., Ciapina L.P.,
   Ciccarelli R.M.B., Coutinho L.L., Cyrino-Santos J.R., El-Dorri H.,
   Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,
   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
   Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
   Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
   Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
   Setubal J.C., Kitajima J.P.;
RA  "Comparison of the genomes of two Xanthomonas pathogens with differing
   host specificities.";
RT  Nature 417:459-463(2002).
RL  Nature 417:459-463(2002).
DR  EMBL; AE012196; AAM40300.1; -
   HSSP; P30749; INVJ.
GO; GO:0006777; P-Mo-molybdopterin cofactor biosynthesis; IEA.

```

```

DR  InterPro; IPR003448; Mb_biosynth_MoAE.
DR  Pfam; PF02391; MoAE; 1.
KW  Complete proteome.
SQ  SEQUENCE 145 AA; 16129 MW; 52518278E41479F1 CRC64;

Query Match      40.3%; Score 50; DB 2; Length 145;
Best Local Similarity 35.3%; Pred. No. 23;
Matches 6; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY  1 WSHPOFEKGGSGWSHPQ 17
   | : | | | | | | | |
DB  127 WKHEHYLEGDAGWLPE 143

Search completed: March 2, 2005, 12:44:22
Job time : 76.3089 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 90.0813 Seconds
(without alignments)

85.869 Million cell updates/sec

Title: β 2010
Perfect score: 124

Sequence: 1 wehpqfekyggawhpfek 20

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : 'A_Geneseq_16dec04':
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20048:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	85.5	36	6	ABP60370 Streptococcus
2	71	57.3	24	6	ABP60362 Streptococcus
3	69	55.6	659	4	AAAB6427 Bili
4	65.5	52.8	35	6	ABP60369 Streptococcus
5	64	51.6	772	8	ADP87577 Human
6	63.5	51.2	845	8	ADP87577 Human
7	62	50.0	396	2	AAW93966 Plasmid
8	62	50.0	396	4	AAW93966 Plasmid
9	62	50.0	396	4	AAW93966 Plasmid
10	62	50.0	400	6	ADA27291 Plasmid
11	62	50.0	400	6	ADA27291 Plasmid
12	62	50.0	400	6	ADA27291 Plasmid
13	62	50.0	400	6	ADA27291 Plasmid
14	62	50.0	400	6	ADA27291 Plasmid
15	60	48.4	646	8	ADP87577 Human
16	60	48.4	858	8	ADP87577 Human
17	59	47.6	236	8	ADP87577 Human
18	59	47.6	236	8	ADP87577 Human
19	58.5	47.2	485	8	ADP87577 Human
20	58.5	47.2	485	8	ADP87577 Human
21	58	46.8	11	6	AAE83873 Bili
22	58	46.8	11	8	ABG74881 Bacteriophage
23	58	46.8	19	6	ABG74881 Bacteriophage
24	58	46.8	19	6	ABG74881 Bacteriophage
25	58	46.8	19	8	ADP87577 Human

ALIGNMENTS

26	58	46.8	19	8	ADP87577 Human	ADP87577 Human
27	58	46.8	183	6	ADA27291 Plasmid	ADA27291 Plasmid
28	58	46.8	183	6	ADA27291 Plasmid	ADA27291 Plasmid
29	58	46.8	209	6	ADA27291 Plasmid	ADA27291 Plasmid
30	58	46.8	258	6	ADA27291 Plasmid	ADA27291 Plasmid
31	58	46.8	258	6	ADA27291 Plasmid	ADA27291 Plasmid
32	58	46.8	539	6	ADP87577 Human	ADP87577 Human
33	58	46.8	539	6	ADP87577 Human	ADP87577 Human
34	58	46.8	661	6	AAE83873 Bili	AAE83873 Bili
35	58	46.8	763	7	ADP87577 Human	ADP87577 Human
36	57	46.0	19	6	ABG74881 Bacteriophage	ABG74881 Bacteriophage
37	57	46.0	19	6	ABG74881 Bacteriophage	ABG74881 Bacteriophage
38	57	46.0	19	6	ABG74881 Bacteriophage	ABG74881 Bacteriophage
39	57	46.0	67	4	AAU47922 Propionibacterium	AAU47922 Propionibacterium
40	57	46.0	67	4	AAU47922 Propionibacterium	AAU47922 Propionibacterium
41	55	44.4	633	6	ADP87577 Human	ADP87577 Human
42	53.5	43.1	267	8	ADR70385 Poliovirus	ADR70385 Poliovirus
43	53.5	43.1	268	8	ADR70385 Poliovirus	ADR70385 Poliovirus
44	53.5	43.1	268	8	ADR70385 Poliovirus	ADR70385 Poliovirus
45	53.5	43.1	268	8	ADR70385 Poliovirus	ADR70385 Poliovirus

RESULT 1
ID ABP60370 standard; peptide: 36 AA.

XX ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

KM Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

XX Key Location/Qualifiers

FT Region 9..28 /note="GGGS repeats 2-5 optionally absent, residues 13-28"

FT Region 9..12 /label="GGGS_repeat"

FT DE10113776-A1.

XX PD 02-OCT-2002.

XX PF 21-MAR-2001; 2001DE-01013776.

XX PF 21-MAR-2001; 2001DE-01013776.

XX PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX PI Schmidt T;

XX DR WPI; 2003-031166/03.

XX PT New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.

XX PS Claim 8; Page 16; 18pp; German.

XX CC The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides

XX CC

CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 CC
 XX

SO Sequence 36 AA;

Query Match 85.5%; Score 106; DB 6; Length 36;
 Best Local Similarity 55.6%; Pred. No. 3.3e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 16; Gaps 1;

QY 1 WSHPOFEK-----GGGSMHPOFEK 20
 |||||
 DB 1 WSHPOFEKGGGGGGGGGGGGGWSHPOFEK 36

RESULT 2

ABP60362
 ID ABP60362 standard; peptide; 24 AA.

XX ABP60362;

XX 28-MAR-2003 (first entry)

XX Streptavidin binding peptide SEQ ID NO 3.

XX Streptavidin; protein chip; microtitre plate; detection.

XX Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 9.16
 /label= unknown

XX DE10113776-A1.

XX 02-OCT-2002.

XX 21-MAR-2001; 2001DE-01013776.

XX 21-MAR-2001; 2001DE-01013776.

XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN:

XX Schmidt T;

XX WPI; 2003-031166/03.

XX New isolated peptide, useful as affinity purification tag for recombinant
 PT protein, comprises at least two high-affinity streptavidin-binding
 PT modules.

XX Disclosure; Page 4; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 CC
 XX

XX Sequence 24 AA;

Query Match 57.3%; Score 71; DB 6; Length 24;
 Best Local Similarity 53.6%; Pred. No. 0.012;
 Matches 15; Conservative 0; Mismatches 1; Indels 12; Gaps 2;

QY 1 WSHPOFEKGGGS-----WSHPOFEK 20
 |||||
 DB 1 WSHPOFE-----KXXXXXXXXXWSHPOFEK 24

RESULT 3

ABP6427
 ID ABP6427 standard; protein; 659 AA.

XX ABP6427;

XX 06-APR-2001 (first entry)

XX Bilin binding-protein associated protein #8.

XX Bilin-binding protein; mutein; BBP; digoxigenin.

XX Unidentified.

XX WO200075308-A1.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-DE001873.

XX 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

XX Skerra A, Schlehuber S;

XX WPI; 2001-071071/08.

XX N-PEDB; AAP25712.

XX New muteins of bilin-binding protein, useful for detecting digoxigenin
 PT being used as label in e.g. binding assays, are very selective for
 PT digoxigenin.

XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
 CC have an amino acid (aa) substitution at at least one of the positions 28,
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
 CC fusion proteins, are used to bind, detect, determine, immobilize or
 CC separate Dig or its conjugates with proteins, nucleic acids,
 CC carbohydrites, other biological or synthetic macromolecules or low
 CC molecular weight compounds, particularly in assays where Dig is being
 CC used as a label. Compared with Dig-specific antibodies, (I) have a
 CC simpler structure and are easier to prepare. They have very high
 CC specificity for Dig, relative to other steroids, and fusion partners may
 CC be attached to either end without compromising their ability to bind
 CC ligand
 CC
 XX

XX Sequence 659 AA;

Query Match 55.6%; Score 69; DB 4; Length 659;
 Best Local Similarity 100.0%; Pred. No. 0.73;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGG 11
 |||||
 DB 198 WSHPOFEKGGG 208

RESULT 4
 ABP60369


```

ID  ABPe0369 standard; peptide; 35 AA.
XX
XX  ABPe0369;
XX
XX  28-MAR-2003 (first entry)
XX
XX  Streptavidin binding peptide SEQ ID NO 10.
XX
XX  Streptavidin; protein chip; microtitre plate; detection.
XX
XX  Synthetic.
XX
XX  Key Location/Qualifiers
FH  Mtec-difference 9..27
FT  /label= unknown
FT  /note= "optionally deleted for 1-15 residues"
XX
XX  DE10113776-A1.
XX
XX  02-OCT-2002.
XX
XX  21-MAR-2001; 2001DE-01013776.
XX
XX  21-MAR-2001; 2001DE-01013776.
XX
XX  (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX  Schmidt T;
XX
XX  WPI; 2003-031166/03.
XX
XX  New isolated peptide, useful as affinity purification tag for recombinant
XX  protein, comprises at least two high-affinity streptavidin-binding
XX  modules.
XX
XX  Claim 7, Page 16; 18pp; German.
XX
XX  The invention relates to an isolated peptide (I) comprising at least two
XX  individual modules separated by 0-50 amino acids, with each containing at
XX  least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
XX  streptavidin binding modules, are useful as affinity handles for
XX  purification of recombinant fusion proteins (FP), also for detecting FP,
XX  e.g. on protein chips or microtitre plates. The modules in (I) bind
XX  strongly to streptavidin, with a co-operative effect that provides
XX  stronger binding than a single tag, but are displaced by a competitor.
XX  (I) does not interfere with the function of attached proteins (II) (so it
XX  may not be essential to remove it); facilitates detection and has easily
XX  controllable binding properties. (I) is particularly used for purifying
XX  FP from dilute solution in batch formats (which use simpler apparatus
XX  than column methods and result in lower loss of FP). The present sequence
XX  is that of a streptavidin binding peptide disclosed with the invention
XX
XX  Sequence 35 AA;
XX
XX  Query Match 52.8%; Score 65.5; DB 6; Length 35;
XX  Best Local Similarity 38.5%; Pred. No. 0.098;
XX  Matches 15; Conservative 0; Mismatches 1; Indels 23; Gaps 2;
XX
XX  1 WSHPOFEKGGG8-----WSHPOFEK 20
XX  |||||
XX  1 WSHPOFE-----KXXXXXXXXXXXXXXXXXWSHPOFEK 35
XX
XX  RESULT 5
XX  ADP87577
XX  ID ADP87577 standard; protein; 772 AA.
XX
XX  AC ADP87577;
XX
XX  DT 23-SEP-2004 (first entry)
XX
XX  DE Human NOVX polypeptide, NOVZF.
XX

```

```

XX  anorectic; antidiabetic; gene therapy; vaccine; obesity; diabetes;
XX  insulin resistance; hybridization probe; chromosome mapping;
XX  tissue typing; preventive medicine; pharmacogenomics; NOVX; human.
XX
XX  Homo sapiens.
XX
XX  WO2004056961-A2.
XX
XX  08-JUL-2004.
XX
XX  27-OCT-2003; 2003WO-US034114.
XX
XX  25-OCT-2002; 2002US-0421239P.
XX  28-OCT-2002; 2002US-0421700P.
XX  31-OCT-2002; 2002US-0422776P.
XX  13-NOV-2002; 2002US-0426197P.
XX  20-DEC-2002; 2002US-0435498P.
XX  20-DEC-2002; 2002US-0435510P.
XX  20-DEC-2002; 2002US-0435568P.
XX  21-MAR-2003; 2003US-0456812P.
XX
XX  (CURA-) CURAGEN CORP.
XX
XX  Berghs C, Catterton E, Ellerman K, Ort T, Rieger D, Chaudhuri A;
XX  WPI; 2004-500293/47.
XX  N-PSDB; ADP87576.
XX
XX  New NOVX nucleic acid molecules and polypeptides useful for preventing or
XX  treating NOVX-associated disorders, e.g. diabetes, insulin resistance or
XX  obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX
XX  Claim 3; SEQ ID NO 32; 570pp; English.
XX
XX  The invention relates to a novel isolated nucleic acid molecule
XX  comprising any of the nucleotide sequences (e.g. 3634, 2127 or 2120 base
XX  pairs) fully defined in the specification; or encodes any of the amino
XX  acid sequences (e.g. 698, 702 or 709 amino acids) fully defined in the
XX  specification. The invention further comprises: an isolated polypeptide
XX  comprising any of the amino acid sequences cited above; a method for
XX  identifying compounds that modulate target polypeptide activity; an
XX  antibody that immunospecifically binds to the target polypeptide, where
XX  the target polypeptide comprises any of the above-mentioned amino acid
XX  sequences, an amino acid sequence that is at least 95% identical to the
XX  above amino acid sequences, an amino acid sequence of at least one domain
XX  of the above-mentioned amino acid sequences, or an amino acid sequence
XX  that is at least 95% identical to the domain of the above amino acid
XX  sequences; a method for identifying a potential therapeutic agent for use
XX  in treatment of a pathology, where the pathology is related to aberrant
XX  expression or aberrant physiological interactions of a target polypeptide
XX  ; and a method of screening for a modulator of activity of or of latency
XX  or predisposition to a pathology associated with a target polypeptide.
XX  The modulating compounds have anorectic and antidiabetic activities. The
XX  nucleic acid sequences of the invention may be used in gene therapy to
XX  treat disorders. The proteins of the invention may be used in creating a
XX  vaccine. The composition and methods are useful for identifying compounds
XX  that modulate protein activity or for diagnosing, preventing or treating
XX  resistance. The nucleic acids are also used as hybridization probes, in
XX  chromosome mapping, tissue typing, preventive medicine, and
XX  pharmacogenomics. This sequence represents a NOVX polypeptide of the
XX  invention.
XX
XX  Sequence 772 AA;
XX
XX  Query Match 51.6%; Score 64; DB 8; Length 772;
XX  Best Local Similarity 68.4%; Pred. No. 4.1;
XX  Matches 13; Conservative 0; Mismatches 4; Indels 2; Gaps 1;
XX
XX  4 POFERKGGG8--WSHPOFEK 20
XX  |||||
XX  754 PAFSHGGSFHNWSHPOFEK 772
XX

```

```

RESULT 6
AD20243
ID AD20243 standard; protein; 845 AA.
AC
XX ADS20243;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human aggrecanase ADAMTS4 mutant E362Q protein with inserted Strep tag.
XX
XX ADAMTS4; a disintegrin-like and metalloprotease;
XX thrombospondin type 1 motif 4; reprotolysin; zinc metalloprotease;
XX aggrecanase; osteopontin; antiinflammatory; antirheumatic; antineumatic;
XX cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;
XX rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
XX human; enzyme; chromosome 1q21-q23; Strep tag; mutant; mutein.
XX
XX Homo sapiens.
XX Synthetic.
XX Unidentified.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 362
XX FT /note= "Wild-type Glu substituted by Gln"
XX FT Misc-difference 520..527
XX FT /note= "Strep tag peptide 2 inserted"
XX
XX WO2004011637-A2.
XX
XX 05-FEB-2004.
XX
XX 29-JUL-2003; 2003WO-US023484.
XX
XX 29-JUL-2002; 2002US-0398721P.
XX
XX (AMAP ) WYETH.
XX (CORC/) CORCORAN C J.
XX (FLAN/) FLANNERY C R.
XX (ZENG/) ZENG W.
XX (RACI/) RACIE L A.
XX (MCDO/) MCDONAGH T.
XX (FREE/) FREEMAN B A.
XX (GEOR/) GEORGIADIS K E.
XX (LAVA/) LAVALLIE E R.
XX
XX Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
XX Freeman BA, Georgiadis KE, Lavallie ER;
XX
XX WPI; 2004-143860/14.
XX
XX New isolated, modified ADAMTS4 (aggrecanase) protein with improved
XX stability useful for identifying inhibitors of the enzyme activity for
XX treating aggrecanase-associated conditions, including osteoarthritis.
XX
XX Claim 9; SEQ ID NO 40; 117pp; English.
XX
XX The invention relates to a novel isolated, modified ADAMTS4 (a
XX disintegrin-like and metalloprotease (repolysin type) with
XX thrombospondin type 1 motif 4) protein with improved stability compared
XX to a naturally occurring, full-length ADAMTS4 protein, where the modified
XX protein differs from the naturally-occurring, full-length ADAMTS4 protein
XX by at least one amino acid. ADAMTS proteins are a subfamily of zinc
XX metalloproteases and include aggrecanases amongst their members. The
XX protein of the invention demonstrates osteopontin, antineumatic,
XX antirheumatic, antineumatic and cytoskeletal activities and may be useful
XX for treating aggrecanase-associated conditions, including osteoarthritis,
XX glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
XX arthritis, periodontal disease and Crohn's disease. The current sequence
XX is that of the human aggrecanase ADAMTS4 mutant E362Q protein with
XX inserted Strep tag of the invention.
XX
XX Sequence, 845 AA;

```

```

Query Match 51.2%; Score 63.5; DB 8; Length 845;
Best Local Similarity 84.6%; Pred. No. 5.3;
Matches 11; Conservative 0; Mismatches 1; Gaps 1;
QY 1 WSHPOFEKGGGSW 13
||| ||| ||| ||| |||
Db 520 WSHPOFEKAGG-W 531
.
RESULT 7
AAW93966
ID AAW93966 standard; protein; 396 AA.
AC
XX AAW93966;
XX
XX 02-JUL-1999 (first entry)
XX
XX Plasmid pBBP20 protein fragment.
XX
XX Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;
XX tissue-specific surface molecule; tumour-specific surface molecule;
XX tumour imaging; bilin binding protein; phage coat protein pIII.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /label= signal_peptide
XX FT Protein 22..396
XX FT /label= fusion_peptide
XX FT /note= "This fusion peptide is constructed from bilin
XX binding protein, a strep-tag II region and a fragment of
XX phage coat protein pIII"
XX FT 22..195
XX FT /note= "Bilin binding protein"
XX FT Region 196..205
XX FT /note= "Strep-Tag II region"
XX FT Misc-difference 206
XX FT /note= "Encoded by TAG"
XX FT Protein 207..396
XX FT /note= "phage coat protein pIII"
XX
XX WO9916873-A1.
XX
XX 08-APR-1999.
XX
XX 25-SEP-1998; 98WO-DE002898.
XX
XX 26-SEP-1997; 97DE-01042706.
XX
XX (SKER/) SKERRA A.
XX
XX Skerra A, Beste G, Schmidt F, Stibora T;
XX
XX WPI; 1999-255093/21.
XX
XX N-PSDB; AAX24099.
XX
XX New anticalins produced by mutation of lipocalin family protein.
XX
XX Disclosure; Page 67-69; 80pp; German.
XX
XX This invention describes the preparation of anticalins from polypeptides
XX of the lipocalin family by mutation of amino acids in the region of the 4
XX peptide loops (at one end of the cylindrical 'leaflet' structure).
XX Anticalins can bind to a predetermined ligand with determinable affinity.
XX Anticalins or fusions of anticalins, are used: (a) when immobilized, for
XX separation; and (b) when labeled, for identification of ligands (or their
XX fusions or conjugates). Typically they bind to tissue- or tumor-specific
XX surface molecules and can be used for tumour imaging or directly for
XX tumour therapy. Mutations can be introduced into lipocalin polypeptides
XX more easily than into antibodies, since lipocalin polypeptides have only
XX 4 (contrast 6) sequence segments. The peptide loops can tolerate amino

```

CC acid changes without significant effect on folding. Anticalins have high
 CC specific affinity, comparable with that for antibodies. This sequence
 CC represents a fusion polypeptide used in the method of the invention
 XX

8Q Sequence 396 AA;

Query Match 50.0%; Score 62; DB 2; Length 396;
 Best Local Similarity 85.7%; Pred. No. 3.8;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGGS 12
 |||||
 Db 198 WSHPOFEKXAGGGS 211

RESULT 8
 AAB46420
 ID AAB46420 standard; protein; 396 AA.
 XX

AC AAB46420;

DT 06-APR-2001 (first entry)

XX Bilin binding-protein associated protein.

XX Bilin-binding protein; muten; BBP; digoxigenin.

XX Unidentified.

XX WO200075308-A1.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-DE001873.

XX 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

XX Skerra A, Schlehuber S;

XX WPI: 2001-071071/08.

DR N-PSDB; AAF25695.

PT New mutenins of bilin-binding protein, useful for detecting digoxigenin
 PT being used as label in e.g. binding assays, are very selective for
 PT digoxigenin;

PS Example 1; Page 55-56; 80pp; German.

XX This invention describes novel polypeptides (I) that are mutenins of bilin
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
 CC have an amino acid (aa) substitution at at least one of the positions 28,
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
 CC fusion proteins, are used to bind, detect, determine, immobilize or
 CC separate Dig or its conjugates with proteins, nucleic acids,
 CC carbohydrates, other biological or synthetic macromolecules or low
 CC molecular weight compounds, particularly in assays where Dig is being
 CC used as a label. Compared with Dig-specific antibodies, (I) have a
 CC simpler structure and are easier to prepare. They have very high
 CC specificity for Dig, relative to other steroids, and fusion partners may
 CC be attached to either end without compromising their ability to bind
 CC ligand

XX Sequence 396 AA;

Query Match 50.0%; Score 62; DB 4; Length 396;
 Best Local Similarity 85.7%; Pred. No. 3.8;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGGS 12
 |||||

Db 198 WSHPOFEKXAGGGS 211

RESULT 9

AAB46423
 ID AAB46423 standard; protein; 396 AA.

XX AAB46423;

DT 06-APR-2001 (first entry)

XX Bilin binding-protein associated protein #4.

XX Bilin-binding protein; muten; BBP; digoxigenin.

XX Unidentified.

XX WO200075308-A1.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-DE001873.

XX 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

XX Skerra A, Schlehuber S;

XX WPI: 2001-071071/08.

DR N-PSDB; AAF25708.

PT New mutenins of bilin-binding protein, useful for detecting digoxigenin
 PT being used as label in e.g. binding assays, are very selective for
 PT digoxigenin.

PS Example 2; Page 62-64; 80pp; German.

XX This invention describes novel polypeptides (I) that are mutenins of bilin
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
 CC have an amino acid (aa) substitution at at least one of the positions 28,
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
 CC fusion proteins, are used to bind, detect, determine, immobilize or
 CC separate Dig or its conjugates with proteins, nucleic acids,
 CC carbohydrates, other biological or synthetic macromolecules or low
 CC molecular weight compounds, particularly in assays where Dig is being
 CC used as a label. Compared with Dig-specific antibodies, (I) have a
 CC simpler structure and are easier to prepare. They have very high
 CC specificity for Dig, relative to other steroids, and fusion partners may
 CC be attached to either end without compromising their ability to bind
 CC ligand

XX Sequence 396 AA;

Query Match 50.0%; Score 62; DB 4; Length 396;
 Best Local Similarity 85.7%; Pred. No. 3.8;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGGS 12
 |||||

Db 198 WSHPOFEKXAGGGS 211

RESULT 10

ADA27291
 ID ADA27291 standard; protein; 400 AA.

XX ADA27291;

DT 20-NOV-2003 (first entry)

DE Plasmid pHNGAL5 fragment protein.

```

XX XX Mucin; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;
KW alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
FH Peptide 1..21
FT /note= "Ompa signal sequence"
FT Protein 22..400
FT /note= "Mature protein; fusion protein of hNGAL, Strep-
FT Tag II and a fragment of phage coat protein pIII"
FT Protein 22..199
FT /note= "Mature modified hNGAL"
FT Peptide 200..209
FT /note= "Strep-Tag II affinity tag"
FT Misc-difference 210
FT /note= "Encoded by TAG"
FT Protein 211..400
FT /note= "Amino acids 217-406 of coat protein pIII"
XX
XX WO2003029462-A1.
XX
XX 10-APR-2003.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlenker S;
XX
XX WPI; 2003-381639/36.
XX
XX N-PSDB; ADA27285.
XX
XX Generating a mutagen of a protein for validating the protein as drug
XX target by subjecting the protein to mutagenesis at sequence positions
XX corresponding to sequence positions of the human neutrophil gelatinase-
XX associated lipocalin.
XX
XX Disclosure; Page 57-58; 68pp; English.
XX
XX The present invention relates to a method for generating mutagens of human
XX neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX subjecting the protein to mutagenesis. In the method of generating a
XX mutagen, a nucleic acid coding for the mutagen of the protein resulting
XX from mutagenesis is operably fused at the 3' end with a gene coding for
XX the coat protein pIII of a filamentous bacteriophage of the M13-family or
XX for a fragment of the coat protein. The present sequence is the protein
XX encoded by a fragment of plasmid pHNGAL5, used to illustrate the
XX invention. This sequence comprises human hNGAL, a Strep-Tag II and a
XX fragment of phage coat protein pIII.
XX
XX Sequence 400 AA;
XX
XX Query Match 50.0%; Score 62; DB 6; Length 400;
XX Best Local Similarity 85.7%; Pred. No. 3.8;
XX Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
OY 1 WSHPOFEK--GGGS 12
DB 202 WSHPOFEKQAGGS 215

```

```

XX XX Plasmid pHNGAL3 fragment protein.
XX DE
XX KW Mucin; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;
XX alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
FH Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..400
FT /note= "Mature protein; fusion protein of hNGAL, Strep-
FT Tag II and a fragment of phage coat protein pIII"
FT Protein 22..199
FT /note= "Mature hNGAL"
FT Peptide 200..209
FT /note= "Strep-Tag II affinity tag"
FT Misc-difference 210
FT /note= "Encoded by TAG"
FT Protein 211..400
FT /note= "Amino acids 217-406 of coat protein pIII"
XX
XX WO2003029462-A1.
XX
XX 10-APR-2003.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlenker S;
XX
XX WPI; 2003-381639/36.
XX
XX N-PSDB; ADA27286.
XX
XX Generating a mutagen of a protein for validating the protein as drug
XX target by subjecting the protein to mutagenesis at sequence positions
XX corresponding to sequence positions of the human neutrophil gelatinase-
XX associated lipocalin.
XX
XX Disclosure; Page 59-61; 68pp; English.
XX
XX The present invention relates to a method for generating mutagens of human
XX neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX subjecting the protein to mutagenesis. In the method of generating a
XX mutagen, a nucleic acid coding for the mutagen of the protein resulting
XX from mutagenesis is operably fused at the 3' end with a gene coding for
XX the coat protein pIII of a filamentous bacteriophage of the M13-family or
XX for a fragment of the coat protein. The present sequence is the protein
XX encoded by a fragment of plasmid pHNGAL3, used to illustrate the
XX invention. This sequence comprises human hNGAL, a Strep-Tag II and a
XX fragment of phage coat protein pIII.
XX
XX Sequence 400 AA;
XX
XX Query Match 50.0%; Score 62; DB 6; Length 400;
XX Best Local Similarity 85.7%; Pred. No. 3.8;
XX Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
OY 1 WSHPOFEK--GGGS 12
DB 202 WSHPOFEKQAGGS 215

```

```

RESULT 11
ID ADA27292 standard; protein; 400 AA.
XX
XX ADA27292;
XX
XX 20-NOV-2003 (first entry)

```

```

RESULT 12
ID ADA00700 standard; protein; 400 AA.
XX
XX ADA00700;
XX
XX ADA00700;

```

```

XX 06-NOV-2003 (first entry)
DT Modified hNGAL related fusion protein SEQ ID NO:13.
XX
DE muclein; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /label= signal
FT /note="modified hNGAL, Strep-tag II and phage coat
FT protein pIII fragment fusion protein"
FT Region 22..400
FT /note="mature hNGAL"
FT Region 200..209
FT /note="Strep-tag II affinity tag"
FT Misc-difference 210
FT /note="amber stop codon"
FT Region 211..400
FT /note="coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX 16-APR-2002; 2002WO-EP004223.
XX
XX (PIR-) PIRIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2003-372000/35.
XX N-PSDB; ADA00731.
XX
XX Generating a muclein of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Example 3; Page 94-95, 122pp; English.
XX
XX The present invention describes a method for generating a muclein of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the muclein has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX mucleins of the protein. Also described: (1) a muclein of hNGAL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the muclein of
XX hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the muclein; (3) a nucleic acid molecule
XX comprising a sequence encoding the muclein of hNGAL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX the muclein of hNGAL, A2m or 24p3 or the fusion protein described above,
XX and a carrier. The mucleins have cytostatic activity, and can be used in
XX gene therapy. The method is useful in generating or producing a muclein of
XX hNGAL, A2m or 24p3 or a fusion protein. The muclein of hNGAL, A2m or 24p3
XX or the fusion protein is useful in detecting a given target by contacting
XX the muclein with a sample suspected of containing the given target under
XX conditions that allow complex formation between the muclein and the given
XX target, and determining the complexed muclein by a suitable signal. The
XX given target is a protein or protein domain, a peptide, a nucleic acid

```

```

CC molecule, an organic molecule or a metal complex and the detection is
CC carried out for validation of the protein as a pharmacological drug
CC target. The muclein may also be used in medicine, e.g. for tumour imaging
CC or directly for cancer therapy. The present sequence represents a
CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion
CC protein given in the exemplification of the present invention.
XX
XX
XX Sequence 400 AA;
XX
XX Query Match 50.0%; Score 62; DB 6; Length 400;
XX Best Local Similarity 85.7%; Pred. No. 3.8;
XX Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
XX
XX Db 1 WSHPOFEK--GGGS 12
XX ||||| |||
XX 202 WSHPOFEKAGGGS 215
XX
XX RESULT 13
XX ADA00707
XX ID ADA00707 standard; protein; 400 AA.
XX AC ADA00707;
XX XX
XX DT 06-NOV-2003 (first entry)
XX
XX DE Modified hNGAL related fusion protein SEQ ID NO:20.
XX
XX muclein; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
XX Synthetic.
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT Protein /label= signal
XX FT /note="modified hNGAL, Strep-tag II and phage coat
XX protein pIII fragment fusion protein"
XX FT Region 22..400
XX FT /note="mature hNGAL"
XX FT Region 200..209
XX FT /note="Strep-tag II affinity tag"
XX FT Misc-difference 210
XX FT /note="amber stop codon"
XX FT Region 211..400
XX FT /note="coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX 16-APR-2002; 2002WO-EP004223.
XX
XX (PIR-) PIRIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2003-372000/35.
XX N-PSDB; ADA00706.
XX
XX Generating a muclein of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Disclosure; Page 104-105, 122pp; English.
XX

```

CC The present invention describes a method for generating a mutein of a
 CC protein selected from a human neutrophil gelatinase-associated lipocalin
 CC (hNGL), rat alpha2-microglobulin-related protein (A2m) and a mouse
 CC 24p3/uterocalin (24p3), where the mutein has a detectable affinity to a
 CC given target, comprising subjecting the protein to mutagenesis at
 CC positions 33-54, 66-83, 94-106 and/or 123-136 of hNGL, resulting in
 CC muteins of the protein. Also described: (1) a mutein of hNGL, A2m or
 CC 24p3 having detectable binding affinity to a given target, obtained by
 CC the method described above; (2) a fusion protein comprising the mutein of
 CC hNGL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
 CC peptide, a signal sequence and/or an affinity tag is operably fused to
 CC the amino or carboxy terminus of the mutein; (3) a nucleic acid molecule
 CC comprising a sequence encoding the mutein of hNGL, A2m or 24p3 or the
 CC fusion protein of (2); and (4) a pharmaceutical composition comprising
 CC the mutein of hNGL, A2m or 24p3 or the fusion protein described above,
 CC and a carrier. The muteins have cytostatic activity, and can be used in
 CC gene therapy. The method is useful in generating or producing a mutein of
 CC hNGL, A2m or 24p3 or a fusion protein. The mutein of hNGL, A2m or 24p3
 CC or the fusion protein is useful in detecting a given target by contacting
 CC the mutein with a sample suspected of containing the given target under
 CC conditions that allow complex formation between the mutein and the given
 CC target, and determining the complexed mutein by a suitable signal. The
 CC given target is a protein or protein domain, a peptide, a nucleic acid
 CC molecule, an organic molecule or a metal complex and the detection is
 CC carried out for validation of the protein as a pharmacological drug
 CC target. The mutein may also be used in medicine, e.g. for tumour imaging
 CC or directly for cancer therapy. The present sequence represents a fusion
 CC modified hNGL, Strep-tag II and phage coat protein pIII fragment fusion
 CC protein given in the exemplification of the present invention.

CC Sequence 400 AA;

Query Match 50.0%; Score 62; DB 6; Length 400;
 Best Local Similarity 85.7%; Pred. No. 3.8;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 WSHPOFEK--GGGS 12
 |||||
 Db 202 WSHPOFEKQAGGS 215

RESULT 14
 ADA00701
 ID ADA00701 standard; protein; 400 AA.
 XX
 AC ADA00701;

XX
 DT 06-NOV-2003 (first entry)
 XX

DE Modified hNGL related fusion protein SEQ ID NO:14.

XX
 KM mutcin; human neutrophil gelatinase-associated lipocalin; hNGL;
 KM rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
 KM 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
 KM tumour imaging; cancer therapy.

XX
 OS Synthetic.
 OS Homo sapiens.
 XX

XX
 FH Key Location/Qualifiers
 FT 1..21
 FT /label= signal
 FT 22..400
 FT Protein

FT /note="modified hNGL, Strep-tag II and phage coat
 FT protein pIII fragment fusion protein"
 FT 22..199
 FT Region

FT /note="mature hNGL"
 FT 200..209
 FT Region

FT /note="Strep-tag II affinity tag"
 FT 210
 FT Misc-difference /note="amber stop codon"
 FT 211..400
 FT Region

FT /note="coat protein pIII fragment 217-406"

XX
 PN WO2003029463-A2.
 XX
 PD 10-APR-2003.
 XX
 PF 18-SEP-2002; 2002WO-EP010490.
 XX
 PR 27-SEP-2001; 2001WO-EP011213.
 XX
 PR 16-APR-2002; 2002WO-EP004223.
 XX
 PA (PIER-) PIERIS PROTEOLAB AG.
 XX
 PI Skerra A, Schlehuber S;
 DR WPI. 2003-372000/35.
 DR N-PSDB; ADN00732.
 XX
 PT Generating a mutein of a protein, e.g. human neutrophil gelatinase-
 associated lipocalin, rat alpha2-microglobulin-related protein or mouse
 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
 PT Disclosure: Page 96-98; 122pp; English.

XX
 CC The present invention describes a method for generating a mutein of a
 CC protein selected from a human neutrophil gelatinase-associated lipocalin
 CC (hNGL), rat alpha2-microglobulin-related protein (A2m) and a mouse
 CC 24p3/uterocalin (24p3), where the mutein has a detectable affinity to a
 CC given target, comprising subjecting the protein to mutagenesis at
 CC positions 33-54, 66-83, 94-106 and/or 123-136 of hNGL, resulting in
 CC muteins of the protein. Also described: (1) a mutein of hNGL, A2m or
 CC 24p3 having detectable binding affinity to a given target, obtained by
 CC the method described above; (2) a fusion protein comprising the mutein of
 CC hNGL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
 CC peptide, a signal sequence and/or an affinity tag is operably fused to
 CC the amino or carboxy terminus of the mutein; (3) a nucleic acid molecule
 CC comprising a sequence encoding the mutein of hNGL, A2m or 24p3 or the
 CC fusion protein of (2); and (4) a pharmaceutical composition comprising
 CC the mutein of hNGL, A2m or 24p3 or the fusion protein described above,
 CC and a carrier. The muteins have cytostatic activity, and can be used in
 CC gene therapy. The method is useful in generating or producing a mutein of
 CC hNGL, A2m or 24p3 or a fusion protein. The mutein of hNGL, A2m or 24p3
 CC or the fusion protein is useful in detecting a given target by contacting
 CC the mutein with a sample suspected of containing the given target under
 CC conditions that allow complex formation between the mutein and the given
 CC target, and determining the complexed mutein by a suitable signal. The
 CC given target is a protein or protein domain, a peptide, a nucleic acid
 CC molecule, an organic molecule or a metal complex and the detection is
 CC carried out for validation of the protein as a pharmacological drug
 CC target. The mutein may also be used in medicine, e.g. for tumour imaging
 CC or directly for cancer therapy. The present sequence represents a fusion
 CC modified hNGL, Strep-tag II and phage coat protein pIII fragment fusion
 CC protein given in the exemplification of the present invention.

XX Sequence 400 AA;

Query Match 50.0%; Score 62; DB 6; Length 400;
 Best Local Similarity 85.7%; Pred. No. 3.8;
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 WSHPOFEK--GGGS 12
 |||||
 Db 202 WSHPOFEKQAGGS 215

RESULT 15
 ADS20251
 ID ADS20251 standard; protein; 646 AA.
 XX
 AC ADS20251;

XX
 DT 18-NOV-2004 (first entry)
 XX

DE Human aggracanease modified ADAMTS4 (MTS4) protein - SEQ ID 49.

XX ADAMTS4: a disintegrin-like and metalloprotease;
KW thrombospondin type 1 motif 4; repolysin; zinc metalloprotease;
KW aggrecanase; osteopontin; antiinflammatory; antirheumatic;
KW cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;
KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
KW human; chromosome 1q21-q23; enzyme.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2004011637-A2.
XX
PD 05-FEB-2004.
XX
PF 29-JUL-2003; 2003WO-US0234484.
XX
PR 29-JUL-2002; 2002US-0398721P.
XX
PA (AMHP) WYETH.
PA (CORC/) CORCORAN C J.
PA (FLAN/) FLANNERY C R.
PA (ZENG/) ZENG W.
PA (RACI/) RACIE L A.
PA (MCDO/) MCDONAGH T.
PA (FREE/) FREEMAN B A.
PA (GEOR/) GEORGIADIS K E.
PA (LAVA/) LAVALLIE E R.
XX
PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
PI Freeman BA, Georgiadis KE, Lavallie ER;
XX
XX WPI: 2004-143860/14.
XX
PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
PT stability useful for identifying inhibitors of the enzyme activity for
PT treating aggrecanase-associated conditions, including osteoarthritis.
XX
PS Claim 9; SEQ ID NO 49; 117pp; English.
XX
CC The invention relates to a novel isolated, modified ADAMTS4 (a
CC disintegrin-like and metalloprotease (repolysin type) with
CC thrombospondin type 1 motif 4) protein with improved stability compared
CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc
CC metalloproteases and include aggrecanases amongst their members. The
CC protein of the invention demonstrates osteopontin, antiinflammatory,
CC antirheumatic, antirheumatic and cytostatic activities and may be useful
CC for treating aggrecanase-associated conditions, including osteoarthritis,
CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
CC arthritis, periodontal disease and Crohn's disease. The current sequence
CC is that of a human aggrecanase modified ADAMTS4 (mTS4) protein of the
CC invention.
XX
SQ Sequence 646 AA;
XX
Query Match 48.4%; Score 60; DB 8; Length 646;
Best Local Similarity 37.5%; Pred. No. 12;
Matches 12; Conservative 2; Mismatches 6; Indels 12; Gaps 1;
OY 1 WSH-----PQFEKGGGWSHPQFEK 20
Db 615 WLHRRRAQLLEILRRPWWGRKGSAMWHPQFEK 646

Search completed: March 2, 2005, 13:02:47
Job time : 91.0813 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 62.1951 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ10
Perfect score: 124
Sequence: 1 webpqfegksgswbhpqfex 20

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	75.0	36	US-10-026-578B-11	Sequence 11, Appl
2	71	57.3	24	US-10-026-578B-3	Sequence 3, Appl
3	65	52.4	36	US-10-026-578B-10	Sequence 10, Appl
4	63.5	51.2	845	US-10-628-432-40	Sequence 40, Appl
5	60	48.4	646	US-10-628-432-49	Sequence 49, Appl
6	60	48.4	858	US-10-628-432-27	Sequence 27, Appl
7	58.5	47.2	485	US-10-628-432-47	Sequence 47, Appl
8	58.5	47.2	697	US-10-628-432-24	Sequence 24, Appl
9	58	46.8	11	US-10-354-983-29	Sequence 29, Appl
10	58	46.8	11	US-10-628-432-25	Sequence 25, Appl
11	58	46.8	661	US-10-354-983-8	Sequence 8, Appl
12	58	46.8	763	US-10-358-283-15	Sequence 15, Appl
13	57	46.0	10	US-10-147-211A-20	Sequence 20, Appl

14	56	45.2	21	9	US-09-809-517A-30	Sequence 30, Appl
15	56	45.2	22	9	US-09-809-517A-33	Sequence 33, Appl
16	56	45.2	24	9	US-09-809-517A-31	Sequence 31, Appl
17	56	45.2	25	9	US-09-809-517A-34	Sequence 34, Appl
18	56	45.2	245	17	US-10-887-228A-1	Sequence 1, Appl
19	56	45.2	246	17	US-10-887-228A-9	Sequence 9, Appl
20	56	45.2	252	17	US-10-887-228A-5	Sequence 5, Appl
21	55	44.4	633	16	US-10-628-432-53	Sequence 53, Appl
22	54	43.5	362	16	US-10-437-963-170711	Sequence 170711, A
23	53.5	43.1	338	15	US-10-425-114-44353	Sequence 44353, A
24	53.5	43.1	837	15	US-10-425-114-53249	Sequence 53249, A
25	53	42.7	180	9	US-09-811-284-249	Sequence 249, App
26	52	41.9	9	9	US-09-983-067-3	Sequence 3, Appl
27	52	41.9	10	9	US-09-809-517A-6	Sequence 6, Appl
28	52	41.9	117	10	US-09-977-137A-4	Sequence 4, Appl
29	52	41.9	117	10	US-09-977-137A-5	Sequence 5, Appl
30	52	41.9	117	10	US-09-977-137A-7	Sequence 7, Appl
31	52	41.9	117	10	US-09-977-137A-8	Sequence 8, Appl
32	52	41.9	117	10	US-09-977-137A-9	Sequence 9, Appl
33	52	41.9	117	10	US-09-977-137A-10	Sequence 10, Appl
34	52	41.9	117	10	US-09-977-137A-11	Sequence 11, Appl
35	52	41.9	117	10	US-09-977-137A-12	Sequence 12, Appl
36	52	41.9	118	10	US-09-977-137A-6	Sequence 6, Appl
37	51	41.1	8	9	US-09-809-517A-9	Sequence 9, Appl
38	51	41.1	8	9	US-09-973-145-7	Sequence 7, Appl
39	51	41.1	8	12	US-09-813-197-8	Sequence 8, Appl
40	51	41.1	8	13	US-10-104-218-5	Sequence 5, Appl
41	51	41.1	8	13	US-10-208-357-9	Sequence 9, Appl
42	51	41.1	8	14	US-10-001-934-8	Sequence 8, Appl
43	51	41.1	8	14	US-10-026-578B-2	Sequence 2, Appl
44	51	41.1	8	14	US-10-026-578B-9	Sequence 9, Appl
45	51	41.1	8	14	US-10-174-368A-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-10-026-578B-11
Sequence 11, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (3)..(28)
OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or mor
OTHER INFORMATION: Repeats may be missing. However, at least one (GGGS) repeat will
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11
Query Match 75.0%; Score 93; DB 14; Length 36;
Best Local Similarity 52.8%; Pred. No. 2e-05;


```
RESULT 5
US-10-628-432-49
; Sequence 49, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 646
; TYPE: PRP
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct E
US-10-628-432-49

Query Match          48.4%; Score 60; DB 16; Length 646;
Best Local Similarity 37.5%; Pred. No. 7.9;
Matches 12; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 WSH-----PQFKGGGSGWSHPQFK 20
DB 615 WLHRRQAQLEILRRPWRGKSGAWSHPQFK 646

RESULT 6
US-10-628-432-27
; Sequence 27, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 858
; TYPE: PRP
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

Query Match          48.4%; Score 60; DB 16; Length 858;
Best Local Similarity 37.5%; Pred. No. 10;
Matches 12; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 WSH-----PQFKGGGSGWSHPQFK 20
DB 827 WLHRRQAQLEILRRPWRGKSGAWSHPQFK 858

RESULT 7
US-10-628-432-47
; Sequence 47, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
```

```
; LENGTH: 485
; TYPE: PRP
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct C
US-10-628-432-47

Query Match          47.2%; Score 58.5; DB 16; Length 485;
Best Local Similarity 40.0%; Pred. No. 9.6;
Matches 12; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

QY 2 SHPQFK-----GGGSGWSHPQFK 20
DB 456 SKKRFDKCMVCGDGGSGGSGAWSHPQFK 485

RESULT 8
US-10-628-432-24
; Sequence 24, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 697
; TYPE: PRP
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Truncated ADAMTS4 molecule
US-10-628-432-24

Query Match          47.2%; Score 58.5; DB 16; Length 697;
Best Local Similarity 40.0%; Pred. No. 13;
Matches 12; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

QY 2 SHPQFK-----GGGSGWSHPQFK 20
DB 668 SKKRFDKCMVCGDGGSGGSGAWSHPQFK 697

RESULT 9
US-10-354-983-29
; Sequence 29, Application US/10354983
; Publication No. US20040044194A1
; GENERAL INFORMATION:
; APPLICANT: AGOSTINO, MICHAEL J.
; APPLICANT: CORCORAN, CHRISTOPHER
; TITLE OF INVENTION: AGGRECANASE MOLECULES
; FILE REFERENCE: 08702.0111-00000
; CURRENT APPLICATION NUMBER: US/10/354,983
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: 60/353,680
; PRIOR FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 11
; TYPE: PRP
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide sequence
US-10-354-983-29

Query Match          46.8%; Score 58; DB 15; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.31;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      10 GGSWSHPQFEK 20
          | : |||||
Db      1  GSAWSHPQFEK 11
```

```

RESULT 10
US-10-628-432-25
; Sequence 25, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:

```

Query Match	46.8%	Score 58	DB 16	Length 11
Best Local Similarity	81.8%	Pred. NO. 0.31		
Matches 9, Conservative	1	Mismatches	1	Indels 0
				Gaps 0

```
QY      10 GGSWSHPQFEK 20
          | : |||||
Db       1 GSAWSHPQFEK 11
```

RESULT 11
 US-10-354-963-8
 Sequence 8, Application US/10354963
 Publication No. US20040004419A1
 GENERAL INFORMATION:
 APPLICANT: AGOSTINO, MICHAEL J.
 APPLICANT: CORCORAN, CHRISTOPHER
 TITLE OF INVENTION: AGGREGANASE MOLECULES
 FILE REFERENCE: 08702.0111-00000
 CURRENT APPLICATION NUMBER: US/10/354,963
 CURRENT FILING DATE: 2003-01-31
 PRIOR APPLICATION NUMBER: 60/353,680
 PRIOR FILING DATE: 2002-01-31
 NUMBER OF SEQ ID NOS: 32
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 8
 LENGTH: 661
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-354-963-8

Query Match	46.8%;	Score 58;	DB 15;	Length 661;
Similarity	81.8%;	Pred.No. 15;		
Best Local	9;	Conservative	1;	Indels 0;
Matches				Gaps 0

```
QY      10 GGSWSHPQFEK 20
          | : |||||
Db      651 GSAWSHPQFEK 661
```

```

RESULT 12
US-10-358-783-15
; Sequence 15, Application US/10358283
; Publication No. US20040054145A1
; GENERAL INFORMATION:
; APPLICANT: WYETH
; TITLE OF INVENTION: TRUNCATED AGGRECANASE MOLECULES
; FILE REFERENCE: 06702-0112-00000
; CURRENT APPLICATION NUMBER: US/10/358,283

```

```

; CURRENT FILING DATE: 2003-02-17
; PRIOR APPLICATION NUMBER: 60/354,592
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
; OS-10-358-283-15

```

Query Match	46.8%;	Score 58;	DB 15;	Length 763;
Best Local Similarity	81.8%;	Pred. No. 17;		
Matches	9;	Conservative	1;	Mismatches 1; Indels 0; Gaps 0;

```
QY      10 GGSWSHPQFEK 20
          | : |||||
Db      753 GSAWSHPQFEK 763
```

```

RESULT 13
US-10-147-211A-20
: Sequence 20, Application US/10147211A
: Publication No. US20030235900A1
: GENERAL INFORMATION:
: APPLICANT: Madison, Edward
: APPLICANT: Yeh, Jiumn-Chern
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
: TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
: FILE REFERENCE: 24745-1616
: CURRENT APPLICATION NUMBER: US/10/147,211A
: CURRENT FILING DATE: 2002-05-14
: PRIOR APPLICATION NUMBER: 60/291,001
: PRIOR FILING DATE: 2001-05-14
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 20
: LENGTH: 10
: TYPE: PRT
: ORGANISM: streptagII sequence with a one glycine linker
US-10-147-211A-20

```

Query Match	46.0%	Score 57;	DB 15;	Length 10;
Best Local Similarity	90.0%	Pred. No. 0.39;		
Matches	9;	Conservative	0;	Mismatches 1;
				Indels 0;
				Gaps 0;

QY	11	GSWSHPQFEK	20
Db	1	GSWSHPQFEK	10

```

RESULT 14
US-09-809-517A-30
Sequence 30, Application US/09809517A
Patent No. US20020034733A1
GENERAL INFORMATION:
APPLICANT: Lohming, Corinna
TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
TITLE OF INVENTION: particles via disulfide bonds
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 30
LENGTH: 21
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:

```

OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match 45.2%; Score 56; DB 9; Length 21;
Best Local Similarity 81.8%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSWSHPQPEK 20
| | | | | | | | | |
DB 11 GAPWSHPQPEK 21

RESULT 15

US-09-809-517A-33
Sequence 33, Application US/09809517A
Patent No. US20020034733A1

GENERAL INFORMATION:

APPLICANT: Lohning, Corinna

TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
TITLE OF INVENTION: Particles via disulfide bonds

FILE REFERENCE: MORPHO/11

CURRENT APPLICATION NUMBER: US/09/809,517A

PRIOR FILING DATE: 2001-03-15

PRIOR APPLICATION NUMBER: EP 99114072.4

PRIOR FILING DATE: 1999-07-20

PRIOR APPLICATION NUMBER: EP 00103551.8

PRIOR FILING DATE: 2000-02-18

NUMBER OF SEQ ID NOS: 41

SOFTWARE: Patentin version 3.0

SEQ ID NO 33

LENGTH: 22

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-33

Query Match 45.2%; Score 56; DB 9; Length 22;

Best Local Similarity 81.8%; Pred. No. 1.1;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSWSHPQPEK 20
| | | | | | | | | |
DB 12 GAPWSHPQPEK 22

Search completed: March 2, 2005, 14:18:52
Job time : 63.1951 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 22.7642 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ10
Perfect score: 124
Sequence: 1 whpghfkgsgswwhpghfek 20

Scoring table: BLOSUM62DX
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/pdata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/pdata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/pdata/1/1aa/6CTUS_COMB.pep:*
6: /cgn2_6/pdata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	45.2	21	4 US-09-809-517A-30	Sequence 30, Appl
2	56	45.2	22	4 US-09-809-517A-31	Sequence 33, Appl
3	56	45.2	24	4 US-09-809-517A-31	Sequence 31, Appl
4	56	45.2	25	4 US-09-809-517A-34	Sequence 34, Appl
5	52	41.9	10	4 US-09-809-517A-6	Sequence 6, Appl
6	52	41.9	11	4 US-09-977-137A-4	Sequence 4, Appl
7	52	41.9	11	4 US-09-977-137A-5	Sequence 5, Appl
8	52	41.9	11	4 US-09-977-137A-7	Sequence 7, Appl
9	52	41.9	11	4 US-09-977-137A-8	Sequence 8, Appl
10	52	41.9	11	4 US-09-977-137A-9	Sequence 9, Appl
11	52	41.9	11	4 US-09-977-137A-10	Sequence 10, Appl
12	52	41.9	11	4 US-09-977-137A-11	Sequence 11, Appl
13	52	41.9	11	4 US-09-977-137A-12	Sequence 12, Appl
14	52	41.9	11	4 US-09-977-137A-6	Sequence 6, Appl
15	51	41.1	8	3 US-08-948-097-2	Sequence 2, Appl
16	51	41.1	8	3 US-09-382-950-7	Sequence 7, Appl
17	51	41.1	8	3 US-09-382-736B-8	Sequence 8, Appl
18	51	41.1	8	4 US-09-619-103-9	Sequence 9, Appl
19	51	41.1	8	4 US-10-104-218-5	Sequence 5, Appl
20	51	41.1	8	4 US-09-809-517A-9	Sequence 9, Appl
21	50	40.3	129	4 US-09-248-796A-22626	Sequence 22626, A
22	50	40.3	662	2 US-08-770-761A-2	Sequence 2, Appl
23	50	40.3	662	2 US-08-770-761A-5	Sequence 5, Appl
24	50	40.3	705	2 US-08-770-761A-7	Sequence 7, Appl
25	48	38.7	155	3 US-08-828-741B-11	Sequence 11, Appl
26	48	38.7	155	3 US-09-160-567-11	Sequence 11, Appl
27	48	38.7	155	4 US-09-710-299-11	Sequence 11, Appl

28	48	38.7	155	4 US-09-509-031-11	Sequence 11, Appl
29	48	38.7	495	3 US-08-828-741B-4	Sequence 4, Appl
30	48	38.7	495	3 US-09-160-567-4	Sequence 4, Appl
31	48	38.7	495	4 US-09-710-299-4	Sequence 4, Appl
32	48	38.7	495	4 US-09-509-031-4	Sequence 4, Appl
33	48	38.7	642	4 US-09-949-016-10283	Sequence 10283, A
34	48	38.7	690	3 US-08-971-089-6	Sequence 10, Appl
35	48	38.7	776	4 US-09-266-225D-10	Sequence 10, Appl
36	48	38.7	777	4 US-09-949-016-6540	Sequence 6540, Ap
37	48	38.7	1858	4 US-09-902-540-12643	Sequence 12643, A
38	47.5	38.3	604	2 US-08-468-576B-12	Sequence 12, Appl
39	47.5	38.3	604	3 US-08-468-579B-12	Sequence 12, Appl
40	47.5	38.3	604	3 US-08-468-577B-12	Sequence 12, Appl
41	47	37.9	228	3 US-09-724-864-46	Sequence 46, Appl
42	47	37.9	259	4 US-09-431-887-33	Sequence 33, Appl
43	47	37.9	434	4 US-09-252-991A-25931	Sequence 25931, A
44	46.5	37.5	1589	3 US-09-356-952-4	Sequence 4, Appl
45	46	37.1	417	4 US-09-134-000C-5002	Sequence 5002, Ap

ALIGNMENTS

```

RESULT 1
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: PARTICLES VIA DISULFIDE BONDS
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRP
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
Query Match          45.2% Score 56; DB 4; Length 21;
Best Local Similarity 81.8%; Pred No. 0.11;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      10 GGSWSHPQFEK 20
      | | | | | | | |
Db       11 GAPWSHPQFEK 21

RESULT 2
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: PARTICLES VIA DISULFIDE BONDS
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41

```

```
SOFTWARE: Patentin version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match          45.2%; Score 56; DB 4; Length 22;
Best Local Similarity 81.8%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      10 GGSWSHPQFEK 20
       | |||||
Db      12 GAPWSHPQFEK 22

RESULT 3
US-09-809-517A-31.
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match          45.2%; Score 56; DB 4; Length 24;
Best Local Similarity 81.8%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      10 GGSWSHPQFEK 20
       | |||||
Db      14 GAPWSHPQFEK 24

RESULT 4
US-09-809-517A-34.
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
```

```
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match          45.2%; Score 56; DB 4; Length 25;
Best Local Similarity 81.8%; Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      10 GGSWSHPQFEK 20
       | |||||
Db      15 GAPWSHPQFEK 25

RESULT 5
US-09-809-517A-6
; Sequence 6, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 6
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-6

Query Match          41.9%; Score 52; DB 4; Length 10;
Best Local Similarity 88.9%; Pred. No. 0.18;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      12 SMSHPQFEK 20
       : |||||
Db      2 AMSHPQFEK 10

RESULT 6
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          41.9%; Score 52; DB 4; Length 117;
Best Local Similarity 88.9%; Pred. No. 2.7;
```


Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
:|||||

DB 109 AMSHPQFEK 117

RESULT 7

US-09-977-137A-5

; Sequence 5, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Cagliat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-5

Query Match Best Local Similarity 41.9%; Score 52; DB 4; Length 117;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
:|||||

DB 109 AMSHPQFEK 117

RESULT 8

US-09-977-137A-7

; Sequence 7, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Cagliat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-7

Query Match Best Local Similarity 41.9%; Score 52; DB 4; Length 117;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
:|||||

DB 109 AMSHPQFEK 117

RESULT 9
US-09-977-137A-8

; Sequence 8, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Cagliat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-8

Query Match Best Local Similarity 41.9%; Score 52; DB 4; Length 117;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
:|||||

DB 109 AMSHPQFEK 117

RESULT 10

US-09-977-137A-9

; Sequence 9, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Cagliat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-9

Query Match Best Local Similarity 41.9%; Score 52; DB 4; Length 117;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
:|||||

DB 109 AMSHPQFEK 117

RESULT 11

US-09-977-137A-10

; Sequence 10, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Cagliat, Jonathan

```
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 117;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
DB 109 AMSHPQFEK 117

RESULT 12
US-09-977-137A-11
; Sequence 11, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 117;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
DB 109 AMSHPQFEK 117

RESULT 13
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
```

```
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 117;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
DB 109 AMSHPQFEK 117

RESULT 14
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 118;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
DB 110 AMSHPQFEK 118

RESULT 15
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skerra, Arne
; TITLE OF INVENTION: Streptavidin Mutelins
; FILE REFERENCE: HUBR 119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; PRIOR FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
```

US-08-948-097-2

Query Match 41.1%; Score 51; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 WSHPOFEK 8
|||
Db 1 WSHPOFEK 8

Search completed: March 2, 2005, 12:25:36
Job time : 23.7642 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 18.3415 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ011

Sequence: 1 webpofekgsgsgsgswahpafek 24

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	61	41.8	99	2 AB2153	RNA binding protei
2	60	41.1	448	1 A56018	transcription fact
3	58	39.7	449	1 S30205	transcription fact
4	58	39.7	451	1 A40168	transcription fact
5	58	39.7	690	2 D98318	ABC protein AGR 1
6	58	39.7	1079	2 F84946	cardamoyl-phosphat
7	57.5	39.4	161	2 B42627	cement precursor p
8	56.5	38.7	346	1 S35500	heterogeneous ribo
9	56	38.4	621	2 T20863	hypothetical prote
10	55	37.7	589	2 S71954	RNA/spDNA-binding
11	54	37.0	333	2 E48423	homeotic protein e
12	54	37.0	447	2 F81678	GTP-binding protei
13	54	37.0	447	2 G71521	probable GTP bindi
14	54	37.0	732	2 T08420	1-phosphatidylinos
15	52	35.6	154	2 E96603	unknown protein F1
16	52	35.6	155	2 T35626	probable membrane
17	52	35.6	425	2 P70038	arabinogalactan en
18	52	35.6	976	1 TVMSMD	macrophage colony-
19	52	35.6	978	2 S16385	macrophage colony-
20	51	34.9	158	2 P96692	hypothetical prote
21	51	34.9	169	1 S38331	glycine-rich RNA-b
22	51	34.9	311	2 T02783	probable homeotic
23	51	34.9	337	2 G84590	probable heat shoc
24	51	34.9	358	2 A54265	CCAAT/enhancer-bin
25	51	34.9	365	2 A26459	helic-destabilizin
26	51	34.9	394	2 C64112	heat shock protein
27	51	34.9	466	2 AH1126	endo-1,4-beta-xyla
28	51	34.9	592	2 E82759	endo-1,4-beta-gluc
29	51	34.9	1240	2 JC5209	insulin receptor s

30	51	34.9	1407	2 B42239	adenylate cyclase
31	50.5	34.6	256	2 UJ0268	major prion protei
32	50.5	34.6	256	2 S37149	prion protein - go
33	50.5	34.6	256	2 A54281	major prion protei
34	50.5	34.6	364	1 QOBB9	BORE1 protein - hu
35	50	34.2	236	1 XMBST9	transcription acti
36	50	34.2	243	2 F90197	hypothetical prote
37	50	34.2	252	2 S53631	major prion protei
38	50	34.2	254	2 A23544	major prion protei
39	50	34.2	282	2 T21222	hypothetical prote
40	50	34.2	288	2 S68798	RNA-binding protei
41	50	34.2	311	2 F86341	hypothetical prote
42	50	34.2	434	2 T01013	hypothetical prote
43	50	34.2	545	2 A38447	oligopeptide ABC t
44	50	34.2	1047	2 G87398	AcirB/AcrD/AcrF fam
45	49.5	33.9	165	2 T24470	hypothetical prote

ALIGNMENTS

RESULT 1
AB2153
RNA binding protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AB2153
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Kanehisa, K.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 6, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2153
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-99 <RUR>
A:Cross-references: UNIPROT:Q9WX37; GB:BA000019; PIDN:BAW74476.1; PID:G17131870; GSPDB:G17131870
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: rbpE
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins
Query Match 41.8%; Score 61; DB 2; Length 99;
Best Local Similarity 64.7%; Pred. No. 0.83;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 SHPOFEKGGSGGSGMS 18
DB 77 ARPREKGRSGGSGMS 93
RESULT 2
A56018
transcription factor Oct-6 - human
N:Alternate names: transcription factor SCIP
C:Species: Homo sapiens (man)
C:Date: 05-May-1995 #sequence_revision 05-May-1995 #text_change 09-Jul-2004
C:Accession: A56018; S30218
R:Faust, I.; Hsu, H.J.; Fuchs, E.
Mol. Cell. Biol. 14, 3263-3275, 1994
A:Title: Oct-6: a regulator of keratinocyte gene expression in stratified squamous epithelium
A:Reference number: A56018; MUID:94217723; PMID:7909356
A:Accession: A56018
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-448 <FAU>
A:Cross-references: UNIPROT:Q03052; GB:L26494; NID:G508989; PIDN:AAA59965.1; PID:G508990
A:Note: it is uncertain whether Met-1 or Met-51 is the initiator
R:Tobler, A.; Schreiber, E.; Fontana, A.
Nucleic Acids Res. 21, 1043, 1993
A:Title: The human Oct-6 POU transcription factor lacks the first 50 amino acids of its
A:Reference number: S30218; MUID:93197134; PMID:8451175

A:Accession: S30218
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 51-448 <TOB>
 A:Cross-references: EMBL:218284; NID:935133; PIDD:CAA79158.1; PID:935134
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
 C:Genetics:
 A:Gene: GDB:POU3F1; OCT-6; OTF6
 A:Cross-references: GDB:138779
 A:Map position: 19pter-19qter
 C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:251-318/Domain: POU domain homology <POU>
 F:337-393/Domain: homeobox homology <HOX>

Query Match 41.1%; Score 60; DB 1; Length 448;
 Best Local Similarity 45.5%; Pred. No. 4.8;
 Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 2 SHPQFKGGGGGGGSHPOFE 23
 Db 67 AHPQWLPTGGGGGDMAGAHLE 88

RESULT 3
 S30205
 transcription factor Oct-6 - mouse
 N:Alternate names: class III POU domain protein SCIP
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Dec-1993 #sequence revision 26-May-1995 #text_change 09-Jul-2004
 C:Accession: S30205; S31226; S13083; S11999
 R:Zimmerman, E.C.; Jones, C.M.; Pet, V.; Hogan, B.L.M.; Magnuson, M.A.
 Nucleic Acid Res. 19, 956, 1991
 A:Title: Nucleotide sequence of mouse SCIP cDNA, a POU-domain transcription factor.
 A:Reference number: S30205; MUID:91204458; PMID:1840678
 A:Accession: S31226
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-449 <ZIM>
 A:Cross-references: UNIPROT:P21952; EMBL:X56959; NID:943681; PIDD:CAA40280.1; PID:943682
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1990
 R:Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.
 Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992
 A:Title: Structure and evolution of four POU domain genes expressed in mouse brain.
 A:Reference number: S31223; MUID:92228768; PMID:1565520
 A:Accession: S31226
 A:Molecule type: DNA
 A:Residues: 1-449 <HAR>
 A:Cross-references: EMBL:W8302; NID:9200450; PIDD:AAA39963.1; PID:9200451
 R:Meijer, D.; Graus, A.; Kraay, R.; Langeveld, A.; Mulder, M.P.; Grosveld, G.
 Nucleic Acid Res. 18, 7357-7365, 1990
 A:Title: The octamer binding factor Oct6: cDNA cloning and expression in early embryonic
 A:Reference number: S13083; MUID:91081324; PMID:1979677
 A:Accession: S13083
 A:Molecule type: mRNA
 A:Residues: 1-449 <MEI>
 A:Cross-references: EMBL:X54628; NID:953505; PIDD:CAA38445.1; PID:953506
 R:Suzuki, N.; Rohdewohld, H.; Neuman, T.; Grunse, F.; Schoeler, H.R.
 EMBO J. 9, 3723-3732, 1990
 A:Title: Oct-6: a POU transcription factor Oct6 expressed in embryonal stem cells and in the
 A:Reference number: S11999; MUID:91006074; PMID:1976514
 A:Accession: S11999
 A:Molecule type: mRNA
 A:Residues: 1-26, 28-449 <SUZ>
 A:Cross-references: EMBL:X57482; NID:953507; PIDD:CAA40720.1; PID:953508
 C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:252-319/Domain: POU domain homology <POU>
 F:338-394/Domain: homeobox homology <HOX>

Query Match 39.7%; Score 58; DB 1; Length 449;
 Best Local Similarity 52.9%; Pred. No. 8.5;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SHPQFKGGGGGGGSHS 18
 Db 67 AHPQWLPTGGGGGDMWA 83

RESULT 4
 A40168
 transcription factor Oct-6 - rat
 N:Alternate names: POU domain protein Tst-1; transcription factor SCIP
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 05-Jun-1992 #sequence revision 05-Jun-1992 #text_change 09-Jul-2004
 C:Accession: A40168; A39694; S05449; S05044
 R:Mouk, E.S.; Kohn, R.; Weinmaster, G.; Trapp, B.D.; Lenke, G.
 Science 249, 1300-1303, 1990
 A:Title: Expression and activity of the POU transcription factor SCIP.
 A:Reference number: A40168; MUID:90378306; PMID:1975954
 A:Accession: A40168
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-451 <MON>
 A:Cross-references: UNIPROT:P20267; GB:M72711; GB:M35205; NID:9206866; PIDD:AAA42118.1;
 R:He, X.; Garretero, R.; Simmons, D.M.; Park, R.E.; Lin, C.R.; Swanson, L.W.; Rosenfeld, M.
 Mol. Cell. Biol. 11, 1739-1744, 1991
 A:Title: Tst-1, a member of the POU domain gene family, binds the promoter of the gene e
 A:Reference number: A39694; MUID:91141528; PMID:1705013
 A:Accession: A39694
 A:Molecule type: mRNA
 A:Residues: 35-451 <HEA>
 A:Cross-references: GB:M63712; NID:9207539; PIDD:AAA42303.1; PID:9207540
 A:Note: the authors translated the codon GCT for residue 249 as pro
 R:He, X.; Treacy, M.N.; Simmons, D.M.; Ingraham, H.A.; Swanson, L.W.; Rosenfeld, M.G.
 Nature 340, 662, 1989
 A:Title: Correction. Expression of a large family of POU-domain regulatory genes in mamm
 A:Reference number: S05449
 A:Accession: S05449
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 265-389 <HEX>
 A:Note: this is a revision to the sequence from reference S05042
 R:He, X.; Treacy, M.N.; Simmons, D.M.; Ingraham, H.A.; Swanson, L.W.; Rosenfeld, M.G.
 Nature 340, 35-42, 1989
 A:Title: Expression of a large family of POU-domain regulatory genes in mammalian brain
 A:Reference number: S05042; MUID:89295573; PMID:2739723
 A:Accession: S05044
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 265-319, 'A', 321-335, 336-389 <HEF>
 A:Note: this sequence has been revised in reference S05449
 C:Genetics:
 A:Gene: tat-1
 C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:254-321/Domain: POU domain homology <POU>
 F:340-396/Domain: homeobox homology <HOX>

Query Match 39.7%; Score 58; DB 1; Length 451;
 Best Local Similarity 52.9%; Pred. No. 8.5;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SHPQFKGGGGGGGSHS 18
 Db 69 AHPQWLPTGGGGGDMWA 85

RESULT 5
 A96318
 ABC protein AGR 1, 3000 (AB015053) (imported) - Agrobacterium tumefaciens (strain C58, Ce
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: D98318
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.

```

Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A87359; MUID:21608551; PMID:11743194
A:Accession: D98318
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-690 <KOR>
A:Cross-references: UNIPROT:Q8UQJ1; GB:AE007870; PIDD:AAK0070.1; PID:g15160053; GSPDB:C
C:Genetics:
A:Gene: AGR_L3000
A:Map position: linear chromosome

Query Match      39.7%; Score 58; DB 2; Length 690;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY      7  EKGGGSGGGSWSHPQFEK 24
      ::||| ||| ||| |||
Db      101 DRGGGGGGGGSHSGFHK 118

RESULT 6
F84946
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain [imported]
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: F84946
R/Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
A:Reference number: A84930; MUID:20445173; PMID:10993077
A:Accession: F84946
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1079 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: carb; BU144
C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin
C:Keywords: ligase

Query Match      39.7%; Score 58; DB 2; Length 1079;
Best Local Similarity 47.6%; Pred. No. 20;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY      4  PQFEKGGSGGGSWSHPQFEK 24
      ||| ||| ||| ||| |||
Db      170 PSFTWGGHGGGIAVNHSEFEK 190

RESULT 7
B42627
cement precursor protein Pc-2 - polychaete (Phragmatopoma californica) (fragments)
C:Species: Phragmatopoma californica
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C:Accession: B42627
R/Walke, J.H.; Jensen, R.A.; Morse, D.E.
Biochemistry 31, 5733-5738, 1992
A:Title: Cement precursor proteins of the reef-building polychaete Phragmatopoma californica
A:Reference number: A42627; MUID:92304953; PMID:1610822
A:Accession: B42627
A:Molecule type: protein
A:Residues: 1-161 <MAI>
A:Cross-references: UNIPROT:Q7M3Q2
C:Comment: This protein becomes a component of the quinine-tanned cement in the tubes bu
F:30,32,53,56,76,89,92,105,108,129,136,149/Modified site: 3',4'-dihydroxyphenylalanine

Query Match      39.4%; Score 57.5; DB 2; Length 161;
Best Local Similarity 30.2%; Pred. No. 3.6;
Matches 13; Conservative 1; Mismatches 10; Indels 19; Gaps 2;

OY      1  WSHPQFEKGGSGG-----GSM-----SHPQFEK 24

```

```

42 WGHFAVHKAGGYGGAHPAAGAGGMAAGGFGGGAHPA VHK 84
Db
RESULT 8
S35500
heterogeneous ribonuclear particle protein homolog - Caenorhabditis elegans
N/Alternate names: heterogeneous nuclear ribonucleoprotein homolog
C/Species: Caenorhabditis elegans
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: S35500; T32620
R/Yasashi, M.; Okumura, K.; Kondo, Y.; Tanaka, T.; Igarashi, H.
Nucleic Acids Res. 20, 4001-4007, 1992
A>Title: cDNA cloning of a novel heterogeneous nuclear ribonucleoprotein gene homologue
A/Reference number: S35500; MUID:92375684; PMID:1354852
A/Accession: S35500
A/Molecule type: mRNA
A/Residues: 1-346 <IWA>
A/Cross-references: UNIPROT:Q22037; EMBL:S43152
R/Du, Z.; Scheer, P.; Andrews, S.
submitted to the EMBL Data Library, December 1997
A/Description: The sequence of C. elegans coamid F42A6.
A/Reference number: Z21201
A/Accession: T32620
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-346 <DUZ>
A/Cross-references: EMBL:AF038613; PIDN:LAB92051.1; GSPDB:GN00022; CESP:F42A6.7
C/Genetics:
A/Experimental source: strain Bristol N2; clone F42A6
A/Gene: CESP:F42A6.7
A/Map position: 4
A/Introns: 9/3; 255/1
C/Superfamily: helix-destabilizing protein; ribonucleoprotein repeat homology
F/24-90/Domain: ribonucleoprotein repeat homology <RMI>
F/115-181/Domain: ribonucleoprotein repeat homology <RM2>
Query Match 38.7%; Score 56.5; DB 1; Length 346;
Best Local Similarity 39.3%; Pred. No. 10;
Matches 11; Conservative 2; Mismatches 8; Indels 7; Gaps 1;
Cy 1 WSHPOFEK-----GGSGGGSWSHPQ 21
Db 271 WGPDPQQGGGGGQGGGGGQGGGGGQ 298
RESULT 9
T20863
hypothetical protein F13H10.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T20863
R/Cottage, A.
submitted to the EMBL Data Library, January 1996
A/Reference number: Z19335
A/Accession: T20863
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-621 <WIL>
A/Cross-references: EMBL:Z68748; PIDN:CAA92953.1; GSPDB:GN00022; CESP:F13H10.3
A/Experimental source: clone F13H10
C/Genetics:
A/Gene: CESP:F13H10.3
A/Map position: 4
A/Introns: 11/3; 45/2; 136/3; 169/2; 222/1; 302/1; 384/1; 417/1; 487/3; 567/2
Query Match 39.4%; Score 56; DB 2; Length 621;
Best Local Similarity 37.5%; Pred. No. 20;
Matches 12; Conservative 4; Mismatches 6; Indels 10; Gaps 1;
Cy 1 WSHPOFEK-----GGSGGGSWSHPQ 22
Db 317 WTIPOWEKQLYDAVSEMEGGGTGGDSWSFDKF 348

```

```

RESULT 10
S71954
RNA/sedna-binding protein TAFI168 - human
N/Alternate names: TATF-binding protein-associated factor TAFI168
C/Species: Homo sapiens (man)
C/Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 09-Jul-2004
C/Accession: S71954
R/Bertolotti, A.; Lutz, Y.; Heard, D.J.; Chambon, P.; Tora, L.
EMBO J. 15, 5022-5031, 1996
A/Title: hTAFI168, a novel RNA/sedna-binding protein with homology to the pro-oncoprotein
A/Reference number: S71954; MUID:97045110; PMID:8890175
A/Accession: S71954
A/Molecule type: mRNA
A/Residues: 1-589 <BER>
A/Cross-references: UNIPROT:Q92804; EMBL:X98893; NID:G1628402; PIDN:CA67398.1; PID:G162
A/Experimental source: HeLa cells
C/Genetics:
A/Gene: taf68
C/Complex: up to 13 TATF-binding protein-associated factors (TAFIs) form together with
C/function:
A/Description: required for activated transcription; binds RNA and single stranded DNA;
C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology
C/Keywords: RNA binding
F/232-307/Domain: ribonucleoprotein repeat homology <RRM>

Query Match          37.0%; Score 54; DB 2; Length 589;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPFKGGGGGGG 15
DB 319 PPFKGGGGGGG 330

RESULT 11
E48423
homeotic protein engrailed 2 - human
N/Alternate names: homeotic protein En-2
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Aug-2004
C/Accession: E48423; A30141
R/Ilogan, C.; Hanks, M.C.; Noble-Topham, S.; Nallathann, D.; Provart, N.J.; Joyner, A.L.
Dev. Genet. 13, 345-358, 1992
A/Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene
A/Reference number: A48423; MUID:93185339; PMID:1363401
A/Accession: E48423
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-333 <LOG>
A/Cross-references: UNIPROT:P19622
R/Poolle, S.U.; Law, M.L.; Kao, F.T.; Lau, Y.F.
Genomics 4, 225-231, 1989
A/Title: Isolation and chromosomal localization of the human En-2 gene.
A/Reference number: A30141; MUID:89233109; PMID:265873
A/Accession: A30141
A/Molecule type: DNA
A/Residues: 230-333 <POO>
A/Genetics:
A/Gene: En-2
C/Superfamily: homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation
F/245-301/Domain: homeobox homology <Hox>

Query Match          37.0%; Score 54; DB 2; Length 333;
Best Local Similarity 58.8%; Pred. No. 19;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 PPFKGGGGGGGWSHP 20
DB 21 PPFKGGGGGGGGSSP 37

```

```

RESULT 12
F81678
GTP-binding protein hflx TC0658 [imported] - Chlamydia muridarum (strain N199)
C/Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: F81678
R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A/Reference number: A81500; MUID:20150255; PMID:10684935
A/Accession: F81678
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-447 <TET>
A/Cross-references: UNIPROT:Q9PK15; GB:AE002334; GB:AE002160; NID:G7190690; PIDN:AAF3948
A/Experimental source: strain N199 (MoPn)
C/Genetics:
A/Gene: TC0658
C/Superfamily: GTP-binding protein hflx; translation elongation factor Tu homology

Query Match          37.0%; Score 54; DB 2; Length 447;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFKGGGGGGG 15
DB 165 WGHLSRQKSGGGGG 179

RESULT 13
G71521
probable GTP binding protein - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C/Species: Chlamydia trachomatis
C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C/Accession: G71521
R/Stephens, R.S.; Kallman, S.; Lammel, C.U.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A/Reference number: A71570; MUID:9900809; PMID:9784136
A/Accession: G71521
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-447 <ARN>
A/Cross-references: UNIPROT:O84383; GB:AE001311; GB:AE001273; NID:G3328799; PIDN:AA6797
A/Experimental source: serotype D, strain UW-3/Cx
C/Genetics:
A/Gene: hflx
C/Superfamily: GTP-binding protein hflx; translation elongation factor Tu homology
C/Keywords: nucleotide binding; P-loop
F/233-240/Region: nucleotide-binding motif A (P-loop)
F/346-349/Region: GTP-binding NKXD motif
F/371-373/Region: GTP-binding SAK/L motif

Query Match          37.0%; Score 54; DB 2; Length 447;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFKGGGGGGG 15
DB 165 WGHLSRQKSGGGGG 179

RESULT 14
T08420
1-phosphatidylinositol 3-kinase (EC 2.7.1.137) - Chlamydomonas reinhardtii (fragment)
N/Alternate names: phosphatidylinositol 3-kinase
C/Species: Chlamydomonas reinhardtii
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C/Accession: T08420
R/Molendijk, A.J.; Irvine, R.F.
Plant Mol. Biol. 37, 53-66, 1998

```


A>Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinosit
 A:Reference number: Z16411; PMID:98281574; PMID:9620264
 A:Accession: T08420
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-732 <MOL>
 A:Cross-references: UNIPROT:004269; EMBL:U97662; NID:g2109288; PIDN:AAC50017.1; PID:g210
 A:Experimental source: strain cw-15
 C:Function:
 A:Description: catalyzes the phosphorylation of phosphoinositides on the 3-hydroxyl grou
 C:Superfamily: phosphatidylinositol 3-kinase, Vps34 type
 C:Keywords: phosphotransferase; signal transduction

Query Match 37.0%; Score 54; DB 2; Length 732;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 HPOFEKGGGGGSGWSHP 20
 | | | | | | | | | |
 DB 379 HSPFGGGGGGGGWPPP 396

RESULT 15

E96603
 unknown protein F14G9.26 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C:Accession: E96603
 R:Theologas, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 aneas, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salizberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; PMID:21016719; PMID:11130712
 A:Accession: E96603
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <STO>
 A:Cross-references: UNIPROT:Q9C7J5; GB:AE005173; NID:g11094725; PIDN:AAG29660.1; GSPDB:G
 C:Genetics:
 A:Gene: F14G9.26
 A:Map position: 1

Query Match 35.6%; Score 52; DB 2; Length 154;
 Best Local Similarity 80.0%; Pred. No. 16;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 9 GGSGGGGWS 18
 | | | | | | | | | |
 DB 115 GGSGGGGWS 124

Search completed: March 2, 2005, 12:28:57
 Job time : 19.3415 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 69.1707 Seconds

(without alignments)
137.824 Million cell updates/sec

Title: SEQ11
Perfect score: 146
Sequence: 1 wshpofekyggsgggswshpofek 24

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	64	43.8	163	2	061UR6
2	62	42.5	429	2	062N21
3	62	42.5	429	2	063XB7
4	62	42.5	704	2	08CUD2
5	61	41.8	99	2	07DL91
6	61	41.8	99	2	09WX37
7	61	41.8	743	2	08CH90
8	61	41.8	929	2	09LGE7
9	60	41.1	448	1	P031_HUMAN
10	60	41.1	454	2	09VHK9
11	60	41.1	455	2	0960C1
12	60	41.1	809	2	07OG12
13	58.5	39.7	82	2	07SKA2
14	58	39.7	150	2	07X423
15	58	39.7	449	1	P031_MOUSE
16	58	39.7	451	1	P031_RAT
17	58	39.7	690	1	07CS58
18	58	39.7	1078	1	CARL_BUCAL
19	57.5	39.4	161	2	07M3Q2
20	57	39.0	1198	2	07S153
21	57	39.0	1515	2	06AV56
22	56.5	38.7	309	2	09S569
23	56.5	38.7	346	2	022037
24	56	38.4	259	1	PRIO_TRIUV
25	56	38.4	434	2	08L4Y9
26	56	38.4	517	2	06ADK6
27	56	38.4	615	1	YSPK_CAEEL
28	56	38.4	617	2	08T3E2
29	56	38.4	640	2	08A120
30	56	38.4	718	2	07XAP4
31	56	38.4	808	2	06BYP9

32	55.5	38.0	285	2	06H7B8	06h7b8 oryza sativ
33	55.5	38.0	420	2	06RR45	06rr45 arabidopsis
34	55	37.7	204	2	09ACP9	09acp9 oryza sativ
35	55	37.7	232	2	07PSQ8	07psq8 anopheles g
36	55	37.7	397	2	095KX8	095kx8 macaca faec
37	55	37.7	501	2	086X94	086x94 homo sapien
38	55	37.7	557	2	08BO46	08bo46 mus musculu
39	55	37.7	592	1	R856_HUMAN	092804 homo sapien
40	55	37.7	959	2	06ZDR8	06zdr8 oryza sativ
41	54.5	37.3	415	2	07PST4	07pst4 anopheles g
42	54	37.0	106	2	08LTI1	08lti1 lycopersico
43	54	37.0	265	2	06H6P3	06h6p3 oryza sativ
44	54	37.0	317	2	094LW4	094lw4 oryza sativ
45	54	37.0	333	1	HME2_HUMAN	p19622 homo sapien

ALIGNMENTS

RESULT 1						
ID	061UR6	PRELIMINARY;	PRT;	163 AA.		
AC	061UR6;					
DT	05-JUL-2004 (TrEMBLrel. 27, Created)					
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)					
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)					
DE	HDCL4384.					
GN	ORFNames=HDCL4384;					
OS	Drosophila melanogaster (Fruit fly).					
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;					
OC	Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;					
OC	Ephydroidea; Drosophilidae; Drosophila.					
OX	NCBI_Taxid=1227;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	PubMed=14709175; DOI=10.1186/gb-2003-5-1-r3;					
RA	Hild M., Beckmann B., Haas S., Koch B., Solovjev V., Busold C.,					
RA	Fellenberg K., Boutros M., Vingron M., Sauer F., Hobeisel J., Pato R,					
RT	"An integrated gene annotation and transpositional profiling approach					
RT	towards the full gene content of the Drosophila genome.";					
RL	Genome Biol. 5:R3-R3(2003).					
CC	-1- MISCELLANEOUS: The sequence shown here is derived from an					
CC	EMBL/GenBank/DBJ third party annotation (TPA) entry.					
DR	EMBL; BK002650; DAA04156.1; -- 002B716DAAABBD39 CRC64;					
SQ	SEQUENCE 163 AA; 18316 MW; 002B716DAAABBD39 CRC64;					
QY	1 WSHPOFEKGGSGGG 15	43.8%;	Score 64;	DB 2;	Length 163;	
DB	77 WPHPOQPRGGGGGGG 91	Best local similarity 66.7%;	Pred. No. 3.8;	Matches 10;	Conservative 1;	Mismatches 4;
					Indels 0;	Gaps 0;
RESULT 2						
ID	062N21	PRELIMINARY;	PRT;	429 AA.		
AC	062N21;					
DT	25-OCT-2004 (TrEMBLrel. 28, Created)					
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)					
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)					
DE	Capuslar polysaccharide biosynthesis/export periplasmic protein.					
GN	ORFNames=BMA0047;					
OS	Burkholderia mallei ATCC 23344.					
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;					
OC	Burkholderiaceae; Burkholderia.					
OX	NCBI_Taxid=243160;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=ATCC 23344;					
RA	Nierman W.C., Deshazer D., Kim H.S., Tettelein H., Nelson K.E.,					
RA	Feldblyum T., Ulrich R.L., Ronning C.M., Brinkac L.M., Daugherty S.C.,					

RA Davidson T.D., Deboy R.T., Dimitrov G., Dodson R.J., Durkin A.S.,
 RA Gwin M.L., Haft D.H., Khouri H., Kolonay J.P., Madupu R.,
 RA Mohammad Y., Nelson W.C., Radune D., Romero C.M., Saria S.,
 RA Selengut J., Shambin C., Sullivan S.A., White C., Yu Y., Zafar N.,
 RA Zhou L., Fraser C.M.,
 RT "Structural flexibility in the Burkholderia mallei genome,"
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251(2004).
 DR EMBL; CP000010; AAU48949.1; -.
 SQ SEQUENCE 429 AA; 45905 MW; 614129D67FBE97 CRC64;

Query Match 42.5%; Score 62; DB 2; Length 429;
 Best Local Similarity 45.0%; Pred. No. 19;
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFKGGSGGSGWSHP 20
 DB 96 WDHPELTRGGNGTGADTSP 115

RESULT 3

AC 063XB7 PRELIMINARY; PRT; 429 AA.
 DT 25-OCT-2004 (TRENBLrel. 28, Created)
 DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
 DE Putative polyaccharide biosynthesis/export protein.
 GN ORFNames=BPSL0619.
 OS Burkholderia pseudomallei K96243.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia.
 OX NCBI_TaxID=272560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K96243;
 RX PubMed=15377794;
 RA Holden M.T.G., Tibball R.W., Peacock S.J., Cerdano-Tarraga A.M.,
 RA Atkins T., Crossman L.C., Platt T., Churcher C., Mungall K.,
 RA Bentley S.D., Sebahia M., Thomson N.R., Baason N., Beacham I.R.,
 RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
 RA Chillingworth T., Cronin A., Crosser B., Davis P., Desnazer D.,
 RA Fellwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagels K.,
 RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
 RA Rabinowitch E., Rutherford K., Sanders M., Simmonds M.,
 RA Songvilay S., Stevens K., Tumapa S., Vesaratchaveest M.,
 RA Whitehead S., Yeates C., Barrett B.G., Oyston P.C.F., Parkhill J.,
 RT "Genomic plasticity of the causative agent of melioidosis,
 RT Burkholderia pseudomallei.".
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
 DR EMBL; BX571965; CAH34612.1; -.
 SQ SEQUENCE 429 AA; 45877 MW; 7A4129C970AECC98 CRC64;

Query Match 42.5%; Score 62; DB 2; Length 429;
 Best Local Similarity 45.0%; Pred. No. 19;
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFKGGSGGSGWSHP 20
 DB 96 WDHPELTRGGNGTGADTSP 115

RESULT 4

AC 08CUD2 PRELIMINARY; PRT; 704 AA.
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DE Guanylyl cyclase alpha 1 subunit.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura I., Suzuki N.,
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB096020; BAC24016.1; -.
 DR HSSP; P30803; IAZS.
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:intracellular signaling cascade; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR01054; G_cyclase.
 DR Pfam; PF00211; Guanylate_cyc; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
 KW lyase.
 SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match 42.5%; Score 62; DB 2; Length 704;
 Best Local Similarity 41.0%; Pred. No. 32;
 Matches 16; Conservative 2; Mismatches 5; Indels 16; Gaps 2;

QY 2 SHPOFK-----GGSG-----GSGWSHPQFK 24
 DB 666 SKPMFQKQADBDGNANFLGKASGVDLVRCGWSHPQFK 704

RESULT 5

AC 07DL91 PRELIMINARY; PRT; 99 AA.
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DE RNA binding protein.
 GN Name=tdpe;
 OS Anabaena variabilis.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 OX NCBI_TaxID=1172;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MS3;
 RX MEDLINE=99216503; PubMed=10198437; DOI=10.1093/nar/27.9.2029;
 RA Maruyama K., Sato N., Ohta N.,
 RT "Conservation of structure and cold-regulation of RNA-binding proteins
 RT in cyanobacteria; probable convergent evolution with eukaryotic
 RT glycine-rich RNA-binding proteins,"
 Nucleic Acids Res. 27:2029-2036(1999).
 RL EMBL; AB003331; BAA77711.1; -.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; RRM_1; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN 1.
 SQ SEQUENCE 99 AA; 10811 MW; 893B196FC8187798 CRC64;

Query Match 41.8%; Score 61; DB 2; Length 99;
 Best Local Similarity 64.7%; Pred. No. 5.1;
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 SHPOFKGGSGGSGWS 18
 DB 77 ARPREKARGSGGSGWS 93

RESULT 6

AC 09WX37 PRELIMINARY; PRT; 99 AA.
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DE RNA binding protein.
 GN Name=tdpe; OrderedLocuNames=112777;

OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxId=103690;
 RN
 ID
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120."
 RL DNA Ref. 8:205-213(2001).
 DR EMBL, AP003590; BAB74476.1; -.
 DR PIR, AB2153; AB2153.
 DR HSSP, P33240; 1P1T.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; RRM_1; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 DR Complete proteome.
 SQ SEQUENCE 99 AA; 10811 MW; 893B196FC8187798 CRC64;

Query Match 41.8%; Score 61; DB 2; Length 99;
 Best Local Similarity 64.7%; Pred. No. 5.1;
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 2 SHPOFEKGGSGGSGWS 18
 Db 77 ARPREKGRSGGSGWS 93

RESULT 7
 OQ8CH90 PRELIMINARY; PRT; 743 AA.
 AC OQ8CH90;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN
 ID
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Nakamura I., Yao Y., Suzuki N.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AB097860; BAC44887.1; -.
 DR HSSP; P30803; 1A2S.
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:lipase activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR001054; G_cyclase.
 DR InterPro; IPR009080; tRNA_Ncy1a_bind.
 DR Pfam; PF00211; Guanylate_cyc; 1.
 DR SMART; SM00444; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASRS_1; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASRS_2; 1.
 DR Lyase.
 SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match 41.8%; Score 61; DB 2; Length 743;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 15 GSWSHPOFEK 24
 Db 734 GSWSHPOFEK 743

RESULT 8
 OQ1GE7 PRELIMINARY; PRT; 929 AA.
 AC OQ1GE7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE P0406H10.1 protein (P0509B06.3 protein).
 GN Name=P0406H10.1; Synonym=P0509B06.3;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
 OC Ehrhacoidae; Oryzaceae; Oryza.
 OX NCBI_TaxId=39947;
 RN
 ID
 RP SEQUENCE FROM N.A.
 RX PubMed=12447438; DOI=10.1038/nature01184;
 RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Karayose Y.,
 RA Wu J., Nimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
 RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayaishi M.,
 RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
 RA Hijiishi S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
 RA Ikono M., Itoh S., Itoh T., Itoh Y., Iwabuuchi A., Kamiya K.,
 RA Karsawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
 RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
 RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
 RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
 RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
 RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yokawa K.,
 RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
 RA Yano M., Jiang J., Gojobori T.,
 RT "The genome sequence and structure of rice chromosome 1."
 RL Nature 420:312-316(2002).
 DR EMBL; AP002524; BAB07977.1; -.
 DR EMBL; AP002903; BAB63468.1; -.
 DR Gramene; OQ1GE7; -.
 DR InterPro; IPR002794; DUF92.
 DR Pfam; PF01940; DUF92; 1.
 DR Pfam; PD014594; DUF92; 1.
 SQ SEQUENCE 929 AA; 100025 MW; F8869AD16F3F6A78 CRC64;

Query Match 41.8%; Score 61; DB 2; Length 929;
 Best Local Similarity 52.9%; Pred. No. 56;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Oy 1 WSHPOFEKGGSGGSGW 17
 Db 537 WSPAMDGGGGGGGVW 553

RESULT 9
 ID P031 HUMAN STANDARD; PRT; 448 AA.
 AC 003052;
 DT 01-JUL-1993 (rel. 26, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 05-JUL-2004 (rel. 44, Last annotation update)
 DE POU domain, class 3, transcription factor 1 (Octamer-binding
 DE transcription factor 6) (Oct-6) (POU-domain transcription factor
 DE SCIP).
 GN Name=POU3F1; Synonym=OCT6, OTF6;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN
 ID
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94217723; PubMed=7909356;
 RA Faus I., Heu H.J., Fuchs E.;
 RT "Oct-6: a regulator of keratinocyte gene expression in stratified
 RT Mol. Cell. Biol. 14:3263-3275(1994)."

```

RN [2]
RP SEQUENCE OF 51-448 FROM N.A.
RX MEDLINE=93197134; PubMed=8451175;
RA Tobler A., Schreiber E., Fontana A.;
RT "The human Oct-6 POU transcription factor lacks the first 50 amino
RL Nucleic Acids Res. 21:1043-1043(1993).
CC -1- FUNCTION: Transcription factor that binds to the octamer motif
CC (5'-ATTTGAT-3'). Thought to be involved in early embryogenesis
CC and neurogenesis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed in embryonal stem cells and in the
CC developing brain.
CC -1- SIMILARITY: Belongs to the POU transcription factor family. Class-
CC 3 subfamily.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; L26494; AAAS9965.1; -
DR EMBL; Z18284; CAA79158.1; -
DR PIR; A56018; A56018.
DR HSSP; P14859; 1COT.
DR TRANSFAC; T00655; -
DR Genew; HGNC:9214; POU3FL.
DR MIM; 602479; -
DR GO; GO:0003700; P:transcription factor activity; TAS.
DR GO; GO:0008366; P:nervous ensheathment; TAS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain like.
DR InterPro; IPR010982; Lambda_like_DNA.
DR InterPro; IPR000327; POU.
DR InterPro; IPR007103; POU_homeo.
DR Pfam; PF00046; Homeobox; 1.
DR Pfam; PF00157; POU; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00028; POUDOMAIN.
DR ProDom; PD000583; POU; 1.
DR ProDom; PD000583; POU; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
DR DNA-binding; Homeobox; Nuclear protein; Transcription regulation.
KW DOMAIN 248 318 POU.
FT DNA_BIND 336 395 Homeobox.
SQ SEQUENCE 448 AA; 45270 MW; 3AA45E029A2C8421 CRC64;
QY 2 SHPOFKGGGGGGGWSHPQFE 23
Db 67 AHPOWLPFGGGGGDWAHALE 88

Query Match 41.1%; Score 60; DB 1; Length 448;
Best Local Similarity 45.5%; Pred. No. 34;
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

```

```

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherier S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abail J.F., Ashby A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottler P.,
RA Buttis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon S., Pollard J., Puri V., Reese M.G.,
RA Palazolo M., Peltzman K., Saunders R.D., Scheeler F., Shen H.,
RA Reinert K., Remington K., Sanders R., Schaefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtka R., Tector C., Turner C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wastarman D.A., Weinstein G.M., Wellenbach J.,
RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhao M., Zhang G., Zhu Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svrtka R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svrtka R.,
RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RT Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

```

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.,
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase/
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RG FlyBase/
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF003754; AAF56525.1; -.
 DR Intact; Q9VBK9; -.
 DR FlyBase; FBgn0039385; CG5913.
 SQ SEQUENCE 454 AA; 49967 MW; 893BA66A5E46CD17 CRC64;

Query Match 41.1%; Score 60; DB 2; Length 454;
 Best Local Similarity 69.2%; Pred. No. 35;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 9 GGGGGGGGWSHPQ 21
 Db 364 GGGGGGGGWNQRPQ 376

RESULT 11

Q960C1 PRELIMINARY; PRT; 455 AA.
 ID 0960C1
 AC 0960C1
 DT 01-DEC-2001 (TRMBLrel. 19, Created)
 DT 01-DEC-2001 (TRMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TRMBLrel. 22, Last annotation update)
 DE SD08037D.
 GN ORFName=CG5913;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunco J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF052130; AK93554.1; -.
 DR FlyBase; FBgn0039385; CG5913.
 SQ SEQUENCE 455 AA; 50095 MW; ED6A9A81669D444D CRC64;

Query Match 41.1%; Score 60; DB 2; Length 455;
 Best Local Similarity 69.2%; Pred. No. 35;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 9 GGGGGGGGWSHPQ 21
 Db 364 GGGGGGGGWNQRPQ 376

RESULT 12

Q70G12 PRELIMINARY; PRT; 809 AA.
 ID 070G12
 AC 070G12
 DT 01-MAR-2004 (TRMBLrel. 26, Created)
 DT 01-MAR-2004 (TRMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
 DE AGCP13861 (Fragment).
 GN Name=agcG51770; ORFName=ENSGANG0000012651;
 OS Anopheles gambiae str. PEST.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 OC NCBI_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

CC EMBL; AAB01008844; EAA05947.1; -.
 DR InterPro; IPR009818; Ataxin-2_C.
 DR InterPro; IPR009604; Ataxin-2_N.
 DR InterPro; IPR010920; Sm_like_fiboprot.
 DR Pfam; PF06741; Ataxin-2_N; 1.
 DR Pfam; PF07145; PAM2; 1.
 FT NON TER 1
 SQ SEQUENCE 809 AA; 86737 MW; 636706D304EFDDBD CRC64;

Query Match 41.1%; Score 60; DB 2; Length 809;
 Best Local Similarity 56.2%; Pred. No. 64;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 7 EKGGGGGGGWSHPQF 22
 Db 243 QHGGGGGGGGMGPQY 258

RESULT 13

Q75KA2 PRELIMINARY; PRT; 82 AA.
 ID 075KA2
 AC 075KA2
 DT 05-JUL-2004 (TRMBLrel. 27, Created)
 DT 05-JUL-2004 (TRMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRMBLrel. 27, Last annotation update)
 DE Hypothetical protein OSJNB0106M04.14.
 GN Name=OSJNB0106M04.14;
 OS Oryza sativa (Japanese cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Buehl C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.L., Teltrin T., Kim M.M., Bera J.U., Jin S.S.,
 RA Padrosh D.W., Tallon L.J., Koo H., Ziemann V., Heiao J., Blunt S.,
 RA Vanden S.S., Riedmuller S.B., Uteback T.T., Feldlyum T.V.,
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.O., Quackenbush J.,
 RA White O., Salzberg S.L., Fraser C.M.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Buehl R.;
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC107207; AAR87333.1; -.
 KW Hypothetical protein.

SQ SEQUENCE 82 AA; 9050 MW; 52B311BC304987C9 CRC64;

Query Match 40.1%; Score 58.5; DB 2; Length 82;
 Best Local Similarity 40.7%; Pred. No. 8.4;
 Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

Qy 4 POFKGGGGGSGS-----WSHPQ 21
 Db 14 PELSGGGGGGGGRRGRRRRRWSRPR 40

RESULT 14
 Q7X243 PRELIMINARY; PRT; 150 AA.
 ID 07X243
 AC 07X243;

```

DR 01-OCT-2003 (Tremblrel. 25, Created)
RT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Curlin-csga protein.
GN Name=csga;
OS Citrobacter sp. Rec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RA Zogal X., Bokantz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515700; CAD56672.1; -.
DR GO: GO:0009289; C:filament; IEA.
DR GO: GO:0007155; P:cell adhesion; IEA.
DR InterPro: IPR009742; Curlin_rpt.
DR Pfam: PF07012; Curlin_rpt; 3.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 39.7%; Score 58; DB 2; Length 150;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 SHPQFEKGGGGSGGSMHPQ 21
Db 22 SVPMQGGGGGGGGSSSGPE 41

RESULT 15
PO3_MOUSE STANDARD; PRT; 449 AA.
ID PO3_MOUSE
AC P21952;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE POU domain, class 3, transcription factor 1 (Octamer-binding
DE transcription factor 6) (Oct-6) (POU-domain transcription factor
DE SCIP).
GN Name=Pou3f1; Synonym=Oct6, Octf-6, Octf6, Scip;
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6 X CBA; TISSUE=Brain;
RC MEDLINE=C57BL/6; PubMed=1979677;
RA Meijer D., Graus A., Kray R., Langeveld A., Mulder M.P., Grosveld G.;
RT "The octamer binding factor Oct6: CDNA cloning and expression in early
RT embryonic cells.";
RL Nucleic Acids Res. 18:7357-7365 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9106074; PubMed=1976514;
RA Suzuki N., Rohdewald H., Neuman T., Gruss P., Schoeler H.R.;
RT "Oct-6: a POU transcription factor expressed in embryonal stem cells
RT and in the developing brain.";
RL EMO J. 9:3723-3732 (1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91204458; PubMed=1840678;
RA Zimmerman E.C., Jones C.M., Fet V., Hogan B.L.M., Magnuson M.A.;
RT "Nucleotide sequence of mouse SCIP cDNA, a POU-domain transcription
RT factor.";
RL Nucleic Acids Res. 19:956-956 (1991).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=92228768; PubMed=1565620;
RA Hara Y., Rovescalli C., Kim Y., Nirenberg M.;

```

```

RT "Structure and evolution of four POU domain genes expressed in mouse
RT brain.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284 (1992).
CC -1- FUNCTION: Transcription factor that binds to the octamer motif
CC (5'-ATTGCAAT-3'). Thought to be involved in early embryogenesis
CC and neurogenesis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed in embryonal stem cells and in the
CC developing brain.
CC -1- SIMILARITY: Belongs to the POU transcription factor family. Class-
CC 3 subfamily.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: X54628; CAA38445.1; -.
DR EMBL: X56959; CAA40280.1; -.
DR EMBL: X57482; CAA40720.1; -.
DR EMBL: M88302; AAA3963.1; -.
DR PIR: S30205; S30205.
DR HSP: P14859; 1COT.
DR TRANSFAC: T00656; -.
DR MGD: MGI:101896; Pou3f1.
DR GO: GO:0005634; C:nucleus; IDA.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR009057; Homeobox_
DR InterPro: IPR010982; Lambda_like_DNA.
DR InterPro: IPR000327; POU.
DR InterPro: IPR007103; POU homeo.
DR Pfam: PF00046; Homeobox; 1.
DR Pfam: PF00157; Pou. 1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00028; POUDOMAIN.
DR ProDom: PD000010; Homeobox; 1.
DR ProDom: PD000583; POU; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00035; POU_2; 1.
DR PROSITE: PS00465; POU_2; 1.
DR DNA-BINDING: Homeobox; Nuclear protein; Transcription regulation.
KW DOMAIN 249 319 POU.
FT DNA BIND 337 396 Homeobox.
FT CONFLICT 35 35 Missing (in Ref. 2).
SQ SEQUENCE 449 AA; 45323 MW; 82DD1335MOC4F43C CRC64;

Query Match 39.7%; Score 58; DB 1; Length 449;
Best Local Similarity 52.9%; Pred. No. 59;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 SHPQFEKGGGGSGGSGMS 18
Db 67 AHPQWLPFGGGGGGMDMA 83

Search completed: March 2, 2005, 12:44:24
Job time: 91.1707 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 108.098 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ11
Perfect score: 146
Sequence: 1 wshpqfkygsgsggswshpqfky 24

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	89.0	36	ABP60370	Abp60370 Streptavi
2	84	57.5	396	AAW93966	AAW93966 Plasmid p
3	84	57.5	396	AAW93966	AAW93966 Plasmid p
4	84	57.5	396	AAW93966	AAW93966 Plasmid p
5	84	57.5	400	ADA27291	ADA27291 Plasmid p
6	84	57.5	400	ADA27291	ADA27291 Plasmid p
7	84	57.5	400	ADA27291	ADA27291 Plasmid p
8	84	57.5	400	ADA27291	ADA27291 Plasmid p
9	84	57.5	400	ADA27291	ADA27291 Plasmid p
10	84	57.5	400	ADA27291	ADA27291 Plasmid p
11	84	57.5	400	ADA27291	ADA27291 Plasmid p
12	84	57.5	400	ADA27291	ADA27291 Plasmid p
13	84	57.5	400	ADA27291	ADA27291 Plasmid p
14	84	57.5	400	ADA27291	ADA27291 Plasmid p
15	84	57.5	400	ADA27291	ADA27291 Plasmid p
16	84	57.5	400	ADA27291	ADA27291 Plasmid p
17	84	57.5	400	ADA27291	ADA27291 Plasmid p
18	84	57.5	400	ADA27291	ADA27291 Plasmid p
19	84	57.5	400	ADA27291	ADA27291 Plasmid p
20	84	57.5	400	ADA27291	ADA27291 Plasmid p
21	84	57.5	400	ADA27291	ADA27291 Plasmid p
22	84	57.5	400	ADA27291	ADA27291 Plasmid p
23	84	57.5	400	ADA27291	ADA27291 Plasmid p
24	84	57.5	400	ADA27291	ADA27291 Plasmid p
25	84	57.5	400	ADA27291	ADA27291 Plasmid p

26	59	40.4	117	5	AAU97557	AAU97557 Synthetic
27	59	40.4	117	5	AAU97555	AAU97555 Synthetic
28	59	40.4	117	5	AAU97556	AAU97556 Synthetic
29	59	40.4	117	5	AAU97552	AAU97552 Synthetic
30	59	40.4	117	5	AAU97560	AAU97560 Synthetic
31	59	40.4	118	5	AAU97554	AAU97554 Synthetic
32	59	40.4	236	8	AD019051	AD019051 Murine an
33	59	40.4	236	8	AD019053	AD019053 Murine an
34	58	39.7	11	6	AAE38373	AAE38373 Eptope t
35	58	39.7	11	8	ADS20228	ADS20228 Strept tag
36	58	39.7	19	6	ABG74881	ABG74881 Bacteriop
37	58	39.7	19	6	ABG74880	ABG74880 Bacteriop
38	58	39.7	19	8	AD139155	AD139155 Construct
39	58	39.7	19	8	AD139156	AD139156 Construct
40	58	39.7	31	8	ADN11208	ADN11208 Peptide m
41	58	39.7	31	8	ADN11225	ADN11225 Peptide m
42	58	39.7	31	8	ADN11207	ADN11207 Peptide m
43	58	39.7	31	8	ADN11226	ADN11226 Peptide m
44	58	39.7	109	3	AAV92150	AAV92150 C19-jun F
45	58	39.7	209	6	ADA00709	ADA00709 Modified

ALIGNMENTS

RESULT 1	ABP60370	standard; peptide; 36 AA.
ID	ABP60370	standard; peptide; 36 AA.
XX	XX	XX
AC	ABP60370;	XX
DT	28-MAR-2003	(first entry)
XX	XX	XX
DE	Streptavidin binding peptide SEQ ID NO 11.	XX
XX	XX	XX
KM	Streptavidin; protein chip; microtitre plate; detection.	XX
XX	XX	XX
OS	Synthetic.	XX
XX	XX	XX
FH	Key	Location/Qualifiers
FT	Region	9..28
FT	Region	/note="GGGS repeats 2-5 optionally absent, residues 13-28"
FT	Region	9..12
FT	Region	/label="GGGS_repeat"
XX	XX	XX
PN	DE10113776-A1.	XX
XX	XX	XX
PD	02-OCT-2002.	XX
XX	XX	XX
PF	21-MAR-2001; 2001DE-01013776.	XX
XX	XX	XX
PR	21-MAR-2001; 2001DE-01013776.	XX
XX	XX	XX
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	XX
XX	XX	XX
PI	Schmidt T;	XX
XX	XX	XX
DR	WPI; 2003-031166/03.	XX
XX	XX	XX
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	XX
PT	modules.	XX
XX	XX	XX
PS	Claim 8; Page 16; 18pp; German.	XX
XX	XX	XX
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I)', which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides	XX

CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 SQ Sequence 36 AA;
 Query Match 89.0%; Score 130; DB 6; Length 36;
 Best Local Similarity 66.7%; Pred. No. 4.2e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 12; Gaps 1;
 QY 1 WSHPOFEK-----GGSGGGGWSHPOFEK 24
 Db 1 WSHPOFEKGGSGGGGGGGGGGGGGWSHPOFEK 36
 RESULT 2
 AAM93966
 XX AAM93966 standard; protein; 396 AA.
 AC AAM93966;
 DT 02-JUL-1999 (first entry)
 XX
 DE Plasmid pBBP20 protein fragment.
 KW Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;
 KM tissue-specific surface molecule; tumour-specific surface molecule;
 XX tumour imaging; bilin binding protein; phage coat protein pIII.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= signal peptide
 FT Protein 22..396
 FT /label= fusion peptide
 FT /note= "This fusion peptide is constructed from bilin
 FT binding protein, a strep-tag II region and a fragment of
 FT phage coat protein pIII"
 FT 22..195
 FT /note= "Bilin binding protein"
 FT Region 196..205
 FT /note= "Strep-Tag II region"
 FT Misc-difference 206
 FT /note= "Encoded by TAG"
 FT Protein 207..396
 FT /note= "phage coat protein pIII"
 XX
 PN WO9916873-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 25-SEP-1998; 98WO-DE002898.
 XX
 PR 26-SEP-1997; 97DE-01042706.
 XX
 PA (SKER/) SKERRA A.
 XX
 PI Skerra A, Beese G, Schmidt F, Stibora T;
 XX WPI, 1999-255093/21.
 DR N-PSDB; AAX24099.
 DR
 XX
 PT New anticalins produced by mutation of lipocalin family protein.
 PS Disclosure; Page 67-69; 80pp; German.
 CC This invention describes the preparation of anticalins from polypeptides
 CC of the lipocalin family by mutation of amino acids in the region of the 4

CC peptide loops (at one end of the cylindrical 'leaflet' structure).
 CC Anticalins can bind to a predetermined ligand with determinable affinity.
 CC Anticalins or fusions of anticalins, are used: (a) when immobilized, for
 CC separation; and (b) when labeled, for identification of ligands (or their
 CC fusions or conjugates). Typically they bind to tissue- or tumour-specific
 CC surface molecules and can be used for tumour imaging or directly for
 CC tumour therapy. Mutations can be introduced into lipocalin polypeptides
 CC more easily than into antibodies, since lipocalin polypeptides have only
 CC 4 (contrast 6) sequence segments. The peptide loops can tolerate amino
 CC acid changes without significant effect on folding. Anticalins have high
 CC specific affinity comparable with that for antibodies. This sequence
 CC represents a fusion polypeptide used in the method of the invention
 SQ Sequence 396 AA;
 Query Match 57.5%; Score 84; DB 2; Length 396;
 Best Local Similarity 88.9%; Pred. No. 0.097;
 Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
 QY 1 WSHPOFEK--GGSGGGGS 16
 Db 198 WSHPOFEKAGGGSGGGGS 215
 RESULT 3
 AAB46420
 XX AAB46420 standard; protein; 396 AA.
 AC AAB46420;
 DT 06-APR-2001 (first entry)
 XX
 DE Bilin binding-protein associated protein.
 KW Bilin-binding protein; mutein; BBP; digoxigenin.
 XX
 OS Unidentified.
 XX
 PN WO200075308-A1.
 XX
 PD 14-DEC-2000.
 XX
 PF 08-JUN-2000; 2000WO-DE001873.
 XX
 PR 08-JUN-1999; 99DE-01026068.
 XX
 PA (SKER/) SKERRA A.
 XX
 PI Skerra A, Schlehuber S;
 XX WPI; 2001-071071/08.
 DR N-PSDB; AAF25695.
 DR
 XX
 PT New muteins of bilin-binding protein, useful for detecting digoxigenin
 FT being used as label in e.g. binding assays, are very selective for
 FT digoxigenin.
 XX
 PS Example 1; Page 55-56; 80pp; German.
 XX
 CC This invention describes novel polypeptides (I) that are muteins of bilin
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
 CC have an amino acid (aa) substitution at at least one of the positions 28,
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
 CC fusion proteins, are used to bind, detect, determine, immobilize or
 CC separate Dig or its conjugates with proteins, nucleic acids,
 CC carbonyldrugs, other biological or synthetic macromolecules or low
 CC molecular weight compounds, particularly in assays where Dig is being
 CC used as a label. Compared with Dig-specific antibodies, (I) have a
 CC simpler structure and are easier to prepare. They have very high
 CC specificity for Dig, relative to other steroids, and fusion partners may
 CC be attached to either end without compromising their ability to bind
 CC ligand

XX Sequence 396 AA;
SQ

Query Match 57.5%; Score 84; DB 4; Length 396;
Best Local Similarity 88.9%; Pred. No. 0.097;
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16
DB 198 WSHPOFEKXAGSGSGGS 215

RESULT 4

AAAB46423
ID AAB46423 standard; protein; 396 AA.

AC AAB46423;

DT 06-APR-2001 (first entry)

DE Bilin binding-protein associated protein #4.

XX Bilin-binding protein; mutcin; BBP; digoxigenin.

XX Unidentified.

OS WO200075308-A1.

XX 14-DEC-2000.

PF 08-JUN-2000; 2000WO-DE001873.

PR 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

PI Skerra A, Schlehuber S;

XX WPI: 2001-071071/08.

DR N-PSDB; AAF25708.

PT New mutcin of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.

XX Example 2; Page 62-64; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin
CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
CC conjugates, do not bind coumarin, testosterone or 4-aminofluorescein, and
CC have an amino acid (aa) substitution at at least one of the positions 28,
CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
CC fusion proteins, are used to bind, detect, determine, immobilize or
CC separate Dig or its conjugates with proteins, nucleic acids,
CC carbohydrates, other biological or synthetic macromolecules or low
CC molecular weight compounds, particularly in assays where Dig is being
CC used as a label. Compared with Dig-specific antibodies, (I) have a
CC simpler structure and are easier to prepare. They have very high
CC specificity for Dig, relative to other steroids, and fusion partners may
CC be attached to either end without compromising their ability to bind
CC ligand

XX Sequence 396 AA;

Query Match 57.5%; Score 84; DB 4; Length 396;
Best Local Similarity 88.9%; Pred. No. 0.097;
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16
DB 198 WSHPOFEKXAGSGSGGS 215

RESULT 5

ADA27291
ID ADA27291 standard; protein; 400 AA.

AC ADA27291;

DT 20-NOV-2003 (first entry)

DE Plasmid pHNGAL5 fragment protein.

XX Mutein; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;
KW alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.

XX Synthetic.

XX Location/Qualifiers

FT Key 1..21
FT Peptide /note="OmpA signal sequence"

FT Protein 22..400
FT /note="Mature protein; fusion protein of hNGAL, Strep-
FT Tag II and a fragment of phage coat protein pIII"

FT Protein 22..199
FT /note="Mature modified hNGAL"

FT Peptide 200..209
FT /note="Strep-Tag II affinity tag"

FT Misc-difference 210
FT /note="Encoded by TNG"

FT Protein 211..400
FT /note="Amino acids 217-406 of coat protein pIII"

XX WO2003029462-A1.

XX 10-APR-2003.

PD 27-SEP-2001; 2001WO-EP011213.

XX 27-SEP-2001; 2001WO-EP011213.

PR (PIER-) PIERIS PROTEOLAB AG.

XX Skerra A, Schlehuber S;

XX WPI: 2003-381639/36.

DR N-PSDB; ADA27285.

PT Generating a mutein of a protein for validating the protein as drug
PT target by subjecting the protein to mutagenesis at sequence positions
PT corresponding to sequence positions of the human neutrophil gelatinase-
PT associated lipocalin.

XX Disclosure; Page 57-58; 68pp; English.

XX The present invention relates to a method for generating muteins of human
CC neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
CC microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
CC subjecting the protein to mutagenesis. In the method of generating a
CC mutein, a nucleic acid coding for the muteins of the protein resulting
CC from mutagenesis is operably fused at the 3' end with a gene coding for
CC the coat protein pIII of a filamentous bacteriophage of the M13-family or
CC for a fragment of the coat protein. The present sequence is the protein
CC encoded by a fragment of plasmid pHNGAL5, used to illustrate the
CC invention. This sequence comprises human hNGAL, a Strep-Tag II and a
CC fragment of phage coat protein pIII.

XX Sequence 400 AA;

Query Match 57.5%; Score 84; DB 6; Length 400;
Best Local Similarity 88.9%; Pred. No. 0.098;
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16
DB 202 WSHPOFEKXAGSGSGGS 219

```

RESULT 6
ID ADA27292-standard; protein; 400 AA.
XX ADA27292;
AC
XX
XX
DT 20-NOV-2003 (first entry)
XX
XX Plasmid pHNGAL3 fragment protein.
XX
XX Mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL; rat;
XX alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX /note="Signal peptide"
XX Protein 22..400
XX /note="Mature protein; fusion protein of hNGAL, Strep-
XX tag II and a fragment of phage coat protein pIII"
XX Peptide 22..199
XX /note="Mature hNGAL"
XX /note="Strep-Tag II affinity tag"
XX Misc-difference 210
XX /note="Encoded by TAG"
XX Protein 211..400
XX /note="Amino acids 217-406 of coat protein pIII"
XX
XX WO2003029462-A1.
XX
XX 10-APR-2003.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2003-381639/36.
XX
XX N-PSDB; ADA27286.
XX
XX
XX PT Generating a mutcin of a protein for validating the protein as drug
XX PT target by subjecting the protein to mutagenesis at sequence positions
XX PT corresponding to sequence positions of the human neutrophil gelatinase-
XX PT associated lipocalin.
XX
XX PS Disclosure; Page 59-61; 68pp; English.
XX
XX
XX The present invention relates to a method for generating mutcins of human
XX neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX subjecting the protein to mutagenesis. In the method of generating a
XX mutcin, a nucleic acid coding for the mutcins of the protein resulting
XX from mutagenesis is operably fused at the 3' end with a gene coding for
XX the coat protein pIII of a filamentous bacteriophage of the M13-family or
XX for a fragment of the coat protein. The present sequence is the protein
XX encoded by a fragment of plasmid pHNGAL3, used to illustrate the
XX invention. This sequence comprises human hNGAL, a Strep-tag II and a
XX fragment of phage coat protein pIII.
XX
XX Sequence 400 AA;
XX
XX Query Match 57.5%; Score 84; DB 6; Length 400;
XX Best Local Similarity 88.9%; Pred. No. 0.098;
XX Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
XX
XX 1 WSHPOFEK--GGSGSGGS 16

```

```

DB
XX ||||| |||||
XX 202 WSHPOFEKAGSGSGGS 219
XX
XX RESULT 7
XX ADA00700
XX ID ADA00700 standard; protein; 400 AA.
XX ADA00700;
AC
XX
XX
XX ADA00700;
XX
XX 06-NOV-2003 (first entry)
XX
XX Modified hNGAL related fusion protein SEQ ID NO:13.
XX
XX Mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
XX Synthetic.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX /label= signal
XX Protein 22..400
XX /note="modified hNGAL, Strep-tag II and phage coat
XX protein pIII fragment fusion protein"
XX Region 22..199
XX /note="mature hNGAL"
XX Region 200..209
XX /note="Strep-tag II affinity tag"
XX Misc-difference 210
XX /note="amber stop codon"
XX Region 211..400
XX /note="coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 16-APR-2002; 2002WO-EP004223.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2003-372000/35.
XX
XX N-PSDB; ADA00731.
XX
XX
XX PT Generating a mutcin of a protein, e.g. human neutrophil gelatinase-
XX PT associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX PT 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX PS Example 3; Page 94-95; 122pp; English.
XX
XX The present invention describes a method for generating a mutcin of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the mutcin has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX mutcins of the protein. Also described: (1) a mutcin of hNGAL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the mutcin of
XX hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the mutcin; (3) a nucleic acid molecule
XX comprising a sequence encoding the mutcin of hNGAL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX

```

CC the mutcin of hNGAL, A2m or 24p3 or the fusion protein described above,
 CC and a carrier. The mutcins have cytostatic activity, and can be used in
 CC gene therapy. The method is useful in generating or producing a mutcin of
 CC hNGAL, A2m or 24p3 or a fusion protein. The mutcin of hNGAL, A2m or 24p3
 CC or the fusion protein is useful in detecting a given target by contacting
 CC the mutcin with a sample suspected of containing the given target under
 CC conditions that allow complex formation between the mutcin and the given
 CC target, and determining the complexed mutcin by a suitable signal. The
 CC given target is a protein or protein domain, a peptide, a nucleic acid
 CC molecule, an organic molecule or a metal complex and the detection is
 CC carried out for validation of the protein as a pharmacological drug
 CC target. The mutcin may also be used in medicine, e.g. for tumour imaging
 CC or directly for cancer therapy. The present sequence represents a
 CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion
 CC protein given in the exemplification of the present invention.

XX SQ Sequence 400 AA;

Query Match 57.5%; Score 84; DB 6; Length 400;
 Best Local Similarity 88.9%; Pred. No. 0.098;
 Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGGGGS 16
 |||||
 Db 202 WSHPOFEKQAGGGSGGS 219

RESULT 8
 ADA00707
 ID ADA00707 standard; protein; 400 AA.

XX AC ADA00707;

XX DT 06-NOV-2003 (first entry)

XX DE Modified hNGAL related fusion protein SEQ ID NO:20.

XX KW mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL;

XX KW rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;

XX KW 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;

XX KW tumour imaging; cancer therapy.

XX OS Synthetic.

XX OS Homo sapiens.

XX XX

XX FH Key

XX FT Peptide

XX FT Protein

XX FT Region

XX FT Region

XX FT Misc-difference

XX FT Region

XX FT Region

XX FT Region

XX FT Region

XX FT Region

XX FT Region

XX FT Region

XX FT Region

DR WPI: 2003-372000/35.
 DR N-PSDB; ADA00706.
 PT Generating a mutcin of a protein, e.g. human neutrophil gelatinase-
 PT associated lipocalin, rat alpha2-microglobulin-related protein or mouse
 PT 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
 PS Disclosure: Page 104-105; 122pp; English.

XX The present invention describes a method for generating a mutcin of a
 CC protein selected from a human neutrophil gelatinase-associated lipocalin
 CC (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
 CC 24p3/uterocalin (24p3), where the mutcin has a detectable affinity to a
 CC given target, comprising subjecting the protein to mutagenesis at
 CC positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
 CC mutcins of the protein. Also described: (1) a mutcin of hNGAL, A2m or
 CC 24p3 having detectable binding affinity to a given target, obtained by
 CC the method described above; (2) a fusion protein comprising the mutcin of
 CC hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
 CC peptide, a signal sequence and/or an affinity tag is operably fused to
 CC the amino or carboxy terminus of the mutcin; (3) a nucleic acid molecule
 CC comprising a sequence encoding the mutcin of hNGAL, A2m or 24p3 or the
 CC fusion protein of (2); and (4) a pharmaceutical composition comprising
 CC the mutcin of hNGAL, A2m or 24p3 or the fusion protein described above,
 CC and a carrier. The mutcins have cytostatic activity, and can be used in
 CC gene therapy. The method is useful in generating or producing a mutcin of
 CC hNGAL, A2m or 24p3 or a fusion protein. The mutcin of hNGAL, A2m or 24p3
 CC or the fusion protein is useful in detecting a given target by contacting
 CC the mutcin with a sample suspected of containing the given target under
 CC conditions that allow complex formation between the mutcin and the given
 CC target, and determining the complexed mutcin by a suitable signal. The
 CC given target is a protein or protein domain, a peptide, a nucleic acid
 CC molecule, an organic molecule or a metal complex and the detection is
 CC carried out for validation of the protein as a pharmacological drug
 CC target. The mutcin may also be used in medicine, e.g. for tumour imaging
 CC or directly for cancer therapy. The present sequence represents a
 CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion
 CC protein given in the exemplification of the present invention.

XX SQ Sequence 400 AA;

Query Match 57.5%; Score 84; DB 6; Length 400;
 Best Local Similarity 88.9%; Pred. No. 0.098;
 Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGGGGS 16
 |||||
 Db 202 WSHPOFEKQAGGGSGGS 219

RESULT 9
 ADA00701
 ID ADA00701 standard; protein; 400 AA.

XX AC ADA00701;

XX DT 06-NOV-2003 (first entry)

XX DE Modified hNGAL related fusion protein SEQ ID NO:14.

XX KW mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL;

XX KW rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;

XX KW 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;

XX KW tumour imaging; cancer therapy.

XX OS Synthetic.

XX OS Homo sapiens.

XX XX

XX FH Key

XX FT Peptide

XX FT Protein

Location/Qualifiers
 1..21
 /label= signal
 22..400
 /note="modified hNGAL, Strep-tag II and phage coat"

```

FT      Region      protein p111 fragment fusion protein"
FT      .199
FT      /note= "mature hNGAL"
FT      Region      200..209
FT      /note= "Strep-tag II affinity tag"
FT      Misc-difference 210
FT      /note= "amber stop codon"
FT      Region      211..400
FT      /note= "coat protein p111 fragment 217-406"
XX      WO2003029463-A2.
XX      10-APR-2003.
XX      18-SEP-2002; 2002WO-EP010490.
XX      27-SEP-2001; 2001WO-EP011213.
XX      16-APR-2002; 2002WO-EP004223.
XX      (PIER-) PIERIS PROTEOLAB AG.
XX      Skerra A., Schlehuber S;
XX      WPI; 2003-372000/35.
XX      N-PSDB; ADA00732.
XX      Generating a mutetin of a protein, e.g. human neutrophil gelatinase-
XX      associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX      24p3/uterocalin, comprising subjecting the protein to mutagenesis.
XX      Disclosure; Page 96-98; 122pp; English.
XX      The present invention describes a method for generating a mutetin of a
XX      protein selected from a human neutrophil gelatinase-associated lipocalin
XX      (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX      24p3/uterocalin (24p3), where the mutetin has a detectable affinity to a
XX      given target, comprising subjecting the protein to mutagenesis at
XX      positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX      mutetins of the protein. Also described: (1) a mutetin of hNGAL, A2m or
XX      24p3 having detectable binding affinity to a given target, obtained by
XX      the method described above; (2) a fusion protein comprising the mutetin of
XX      hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX      peptide, a signal sequence and/or an affinity tag is operably fused to
XX      the amino or carboxy terminus of the mutetin; (3) a nucleic acid molecule
XX      comprising a sequence encoding the mutetin of hNGAL, A2m or 24p3 or the
XX      fusion protein of (2); and (4) a pharmaceutical composition comprising
XX      the mutetin of hNGAL, A2m or 24p3 or a fusion protein described above,
XX      and a carrier. The mutetins have cytostatic activity, and can be used in
XX      gene therapy. The method is useful in generating or producing a mutetin of
XX      hNGAL, A2m or 24p3 or a fusion protein. The mutetin of hNGAL, A2m or 24p3
XX      or the fusion protein is useful in detecting a given target by contacting
XX      the mutetin with a sample suspected of containing the given target and the given
XX      conditions that allow complex formation between the mutetin and the target.
XX      The target, and determining the complexed mutetin by a suitable signal. The
XX      given target is a protein or protein domain, a peptide, a nucleic acid
XX      molecule, an organic molecule or a metal complex and the detection is
XX      carried out for validation of the protein as a pharmacological drug
XX      target. The mutetin may also be used in medicine, e.g. for tumour imaging
XX      or directly for cancer therapy. The present sequence represents a
XX      modified hNGAL, Strep-tag II and phase coat protein p111 fragment fusion
XX      protein given in the exemplification of the present invention.
XX      Sequence 400 AA;
XX
XX      Query Match      57.5%; Score 84; DB 6; Length 400;
XX      Best Local Similarity 88.3%; Pred. No. 0.098;
XX      Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
XX
XX      1 WSHPOFEK--GGSGGGGS 16
XX      |||||
XX      202 WSHPOFEKQAGSGSGGS 219

```

```

RESULT 10
AAB46427
ID      AAB46427 standard; protein; 659 AA.
XX      AAB46427;
XX      06-APR-2001 (first entry)
XX      Bilin binding-protein associated protein #8.
XX      Bilin-binding protein; mutetin; BBP; digoxigenin.
XX      Unidentified.
XX      WO200075308-A1.
XX      14-DEC-2000.
XX      08-JUN-2000; 2000WO-DE001873.
XX      08-JUN-1999; 99DE-01026068.
XX      (SKER/) SKERRA A.
XX      Skerra A., Schlehuber S;
XX      WPI; 2001-071071/08.
XX      N-PSDB; AAF25712.
XX      New mutetins of bilin-binding protein, useful for detecting digoxigenin
XX      being used as label in e.g. binding assays, are very selective for
XX      digoxigenin.
XX      Example 4; Page 72-74; 80pp; German.
XX      This invention describes novel polypeptides (I) that are mutetins of bilin
XX      -binding protein (BBP), which can bind digoxigenin (Dig) or its
XX      conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
XX      have an amino acid (aa) substitution at at least one of the positions 28,
XX      31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127 (I), or its
XX      fusion proteins, are used to bind, detect, determine, immobilize or
XX      separate Dig or its conjugates with proteins, nucleic acids,
XX      carbohydrates, other biological or synthetic macromolecules or low
XX      molecular weight compounds, particularly in assays where Dig is being
XX      used as a label. Compared with Dig-specific antibodies, (I) have a
XX      simpler structure and are easier to prepare. They have very high
XX      specificity for Dig, relative to other steroids, and fusion partners may
XX      be attached to either end without compromising their ability to bind
XX      ligand
XX      Sequence 659 AA;
XX
XX      Query Match      51.4%; Score 75; DB 4; Length 659;
XX      Best Local Similarity 92.3%; Pred. No. 1.8;
XX      Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX      1 WSHPOFEKGGSGS 13
XX      |||||
XX      198 WSHPOFEKGGSGG 210

```

RESULT 11

```

AD820249
ID      AD820249 standard; protein; 485 AA.
XX      AD820249;
XX      18-NOV-2004 (first entry)
XX      Furin-processed human aggrecanase ADAMTS4 truncated protein w Strep tag.
XX      ADAMTS4, a disintegrin-like and metalloprotease;
XX      thrombospondin type 1 motif 4; repolysin; zinc metalloprotease;
XX

```

	KM	rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease; human; chromosome 1q21-q23; enzyme; truncation; mature; furin cleavage site.
	XK	Homo sapiens.
	OS	undetermined.
	XX	Key Location/Qualifiers
	FH	Misc-difference 475..485 /note= "Residues corresponding to positions 687-837 in the wild-type replaced by Strep tag"
	FT	
	FT	
	PX	MO2004011637-A2.
	PN	
	PD	05-FEB-2004.
	XX	
	PF	29-JUL-2003; 2003WO-US023464.
	XX	
	PR	29-JUL-2002; 2002US-0398721P.
	XX	
	PA	(AMHP) WYETH.
	PA	(CORC/) CORCORAN C J.
	PA	(PLAN/) FLANNERY C R.
	PA	(ZENQ/) ZENG W.
	PA	(RACI/) RACIE L A.
	PA	(MCDO/) MCDONAGH T.
	PA	(FREE/) FREEMAN B A.
	PA	(GEOR/) GEORGADIS K E.
	PA	(LAVA/) LAVALLIE E R.
	XX	
	P1	Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
	P1	Freeman BA, Georgiadis KE, Lavallie ER;
	DR	WPI; 2004-143860/14.
	XX	
	PT	New isolated, modified ADAMTS4 (aggrecanase) protein with improved stability useful for identifying inhibitors of the enzyme activity for creating aggrecanase-associated conditions, including osteoarthritis.
	PS	Claim 9; SEQ ID NO 47; 117pp; English.
	XX	The invention relates to a novel isolated, modified ADAMTS4 (a chondroitinsulfate-like and metalloproteinase (reprolysin type) with disintegrin-like motif 4) protein with improved stability compared to a naturally occurring, full-length ADAMTS4 protein, where the modified protein differs from the naturally-occurring, full-length ADAMTS4 protein by at least one amino acid. ADAMTS proteins are a subfamily of zinc metalloproteinases and include aggrecanases amongst their members. The protein of the invention demonstrates osteopathic, anti-inflammatory, antirheumatic, antineuritic and cytostatic activities and may be useful for treating aggrecanase-associated conditions, including osteoarthritis, glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic arthritis, periodontal disease and Crohn's disease. The current sequence is that of the furin-processed human aggrecanase ADAMTS4 truncated protein with Strep tag of the invention.
	SO	Sequence 485 AA:
Oy		Query Match 49.0%; Score 71.5; DB 8; Length 485; Best Local Similarity 46.7%; Pred. No. 3.3; Matches 14; Conservative 3; Mismatches 6; Indels 7; Gaps 1 2 SHPOFEK-----GGSGGCGSMSHPOFEK 24 :: : Db 456 SKKKFDCMKVCCGDGSCGSASMSHPOFEK 485
RESULT 12		
AIDS020227		
ID AIDS020227		standard; protein; 697 AA.

AC	AD520227;
XX	
DT	18-NOV-2004 (first entry)
DE	Human aggrecanase ADAMTS4 truncated protein with Strep tag - SEQ ID 24.
XX	
KW	ADAMTS4; a disintegrin-like and metalloprotease; thrombospondin type 1 motif 4; reprotysin; zinc metalloprotease; aggrecanase; osteoarthritic; antiinflammatory; antiarthritic; antirheumatic; cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint; rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease; human; enzyme; chromosome 1q21-q23; truncation; Strep tag.
XX	
OS	Homo sapiens.
OS	Unidentified.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 687..697
FT	/note= "Wild-type residues 687-837 replaced by Strep tag"
FN	MO2004011637-A2.
XX	
PD	05-FEB-2004.
XX	
PE	29-JUL-2003; 2003WO-US023484.
PR	29-JUL-2002; 2002US-0398721P.
XX	
PA	(AMHP) WYETH.
PA	(CORC/) CORCORAN C J.
PA	(PLAN/) FLANNERY C R.
PA	(ZENG/) ZENG W.
PA	(RACI/) RACIE L A.
PA	(MCDO/) MCDONAGH T.
PA	(FREE/) FREEMAN B A.
PA	(GEOR/) GEORGIADIS K E.
PA	(LAVN/) LAVALLIE E R.
P1	Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
P1	Freeman BA, Georgiadis KE, Lavallie ER;
DR	WPI; 2004-143860/14.
XX	
XX	New isolated, modified ADAMTS4 (aggrecanase) protein with improved stability useful for identifying inhibitors of the enzyme activity for treating aggrecanase-associated conditions, including osteoarthritis.
PS	Claim 9; SEQ ID NO 24; 117pp; English.
XX	
CC	The invention relates to a novel isolated, modified ADAMTS4 (a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif 4) protein with improved stability compared to a naturally occurring, full-length ADAMTS4 protein, where the modified protein differs from the naturally-occurring, full-length ADAMTS4 protein by at least one amino acid. ADAMTS proteins are a subfamily of zinc metalloproteases and include aggrecanases amongst their members. The protein of the invention demonstrates osteopathic, antiinflammatory, antiarthritic, antirheumatic and cytoskeletal activities and may be useful for treating aggrecanase-associated conditions, including osteoarthritis, glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic arthritis, periodontal disease and Crohn's disease. The current sequence is that of the human aggrecanase ADAMTS4 truncated protein with Strep tag of the invention.
XX	
SQ	Sequence 697 AA;
Query Match	49.0%; Score 71.5; DB 8; Length 697;
Best Local Similarity	46.7%; Pred. No. 4.8;
Matches 14; Conservative 3; Mismatches 6; Indels 7; Gaps 1	
2 SHPOFEK-----CGSGGCGSWSHPOFEK 24	
: : : : : : : : : : : : : : : : : :	
668 SKKKFDKCMVCGGDSSGCGSAWSHPOFEK 697	

RESULT 13
 ADS20243
 ID ADS20243 standard; protein; 845 AA.
 XX
 AC ADS20243;
 XX
 DT 18-NOV-2004 (first entry)
 DE Human aggrecanase ADAMTS4 mutant E362Q protein with inserted Strep tag.
 XX
 KW ADAMTS4, a disintegrin-like and metalloprotease;
 XX thrombospondin type 1 motif 4; reprolysin; zinc metalloprotease;
 KW aggrecanase; osteopontin; antiinflammatory; antiarthritic; antineumatic;
 KW cytosolic; osteoarthritis; glioma; cancer; inflammatory joint;
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
 KW human; enzyme; chromosome 1q21-q23; Strep tag; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 362
 FT /note="Wild-type Glu substituted by Gln"
 FT Misc-difference 520..527
 FT /note="Strep tag peptide 2 inserted"
 FT
 PN WO2004011637-A2.
 XX
 PD 05-FEB-2004.
 XX
 PF 29-JUL-2003; 2003WO-US023484.
 XX
 PR 29-JUL-2002; 2002US-0398721P.
 XX
 PA (AMHP) WYETH.
 PA (CORC/) CORCORAN C J.
 PA (FLAN/) FLANNERY C R.
 PA (ZENG/) ZENG W.
 PA (RACI/) RACIE L A.
 PA (MCDO/) MCDONAGH T.
 PA (FRED/) FREEMAN B A.
 PA (GEOR/) GEORGIADIS K E.
 PA (LAVA/) LAVALLIE E R.
 XX
 PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
 PI Freeman BA, Georgiadis KE, Lavallie ER;
 XX
 DR WPI; 2004-143860/14.
 XX
 PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
 PT stability useful for identifying inhibitors of the enzyme activity for
 PT treating aggrecanase-associated conditions, including osteoarthritis.
 XX
 PS Claim 9; SEQ ID NO 40; 117pp; English.
 XX
 CC The invention relates to a novel isolated, modified ADAMTS4 (a
 CC disintegrin-like and metalloprotease (reprolysin type) with
 CC thrombospondin type 1 motif 4) protein with improved stability compared
 CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
 CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
 CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc
 CC metalloproteases and include aggrecanases amongst their members. The
 CC protein of the invention demonstrates osteopontin, antiinflammatory,
 CC antiarthritic, antineumatic and cytostatic activities and may be useful
 CC for treating aggrecanase-associated conditions, including osteoarthritis,
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
 CC arthritis, periodontal disease and Crohn's disease. The current sequence
 CC is that of the human aggrecanase ADAMTS4 mutant E362Q protein with
 CC inserted Strep tag of the invention.
 XX

SEQ Sequence 845 AA;
 Query Match 47.6%; Score 69.5; DB 8; Length 845;
 Best Local Similarity 72.2%; Pred. No. 9.8;
 Matches 13; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 QY 1 WSHPOFEKGGSG-GGSW 17
 |||||
 DB 520 WSHPOFEKAGGWGMPW 537
 |||||
 RESULT 14
 ID ABP60362 standard; peptide; 24 AA.
 XX
 AC ABP60362;
 XX
 DT 28-MAR-2003 (first entry)
 DE Streptavidin binding peptide SEQ ID NO 3.
 XX
 KW Streptavidin; protein chip; microtitre plate; detection.
 XX
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 9..16
 FT /label= unknown
 FT
 PN DE10113776-A1.
 XX
 PD 02-OCT-2002.
 XX
 PF 21-MAR-2001; 2001DE-01013776.
 XX
 PR 21-MAR-2001; 2001DE-01013776.
 XX
 PA (BIOA-) INST BIOANALYTIC GMBH GOETTINGEN.
 XX
 PI Schmidt T;
 XX
 DR WPI; 2003-031166/03.
 XX
 PT New isolated peptide, useful as affinity purification tag for recombinant
 PT protein, comprises at least two high-affinity streptavidin-binding
 PT modules.
 XX
 PS Disclosure; Page 4; 18pp; German.
 XX
 CC The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX
 SQ Sequence 24 AA;
 Query Match 47.3%; Score 69; DB 6; Length 24;
 Best Local Similarity 46.9%; Pred. No. 0.33;
 Matches 15; Conservative 0; Mismatches 1; Indels 16; Gaps 2;
 QY 1 WSHPOFEKGGSGGGS-----WSHPOFEK 24
 |||||
 DB 1 WSHPOFE-----KXXXXXXXXXWSHPOFEK 24

RESULT 15

AAO19986 standard; protein; 42 AA.

AAO19986;

24-MAY-2002 (first entry)

Protein of (double hexahistidine) - flexible linker upstream.

Protein array; cell free system; gridded format; covalent; non-covalent; ligand; phage display library; cellular protein expression profile; ribosome display library; post-translational modification; hexahistidine; flexible linker.

Unidentified.

WO200214860-A1.

21-FEB-2002.

15-AUG-2001; 2001WO-GB003657.

15-AUG-2000; 2000GB-00020016.

01-DEC-2000; 2000GB-00029309.

16-MAR-2001; 2001GB-00006610.

07-JUN-2001; 2001GB-00013883.

14-JUL-2001; 2001GB-00017232.

(DISC-) DISCERNA LTD.

He M, Tauesig MJ;

WPI; 2002-241989/29.

N-PSDB; AAK98912.

Protein arrays useful for identifying interactions of arrayed proteins

with other molecules, in which proteins are produced by in vitro

synthesis using cell-free transcription and translation systems.

Disclosure; Fig 3B; 61pp; English.

The invention relates to a method for producing a protein array made by

transcription and translation of DNA, or by translation of mRNA, by a

cell free system in vitro to produce individual proteins, domains or

peptides distributed in a gridded format. The peptides of the invention

contain amino acid sequences enabling covalent/non-covalent attachment to

a surface/bead, such that the peptides can be arranged in the grid format

after interaction with suitable ligands or reagents on the surface. The

method of the invention is useful for identifying interactions of the

arrayed proteins with other molecules, including antibodies, other

proteins or domains, peptides, small ligands, cell extracts and nucleic

acids. The protein array is also useful for identifying interactions of

arrayed proteins with other molecules displayed in a library, such as a

phage display library or ribosome display library. The protein array is

useful for studying cellular protein expression profiles and post-

translational modification of cellular proteins. This sequence of the

invention represents a protein of a double hexahistidine with a flexible

linker sequence upstream

Sequence 42 AA:

Query Match 45.2%; Score 66; DB 5; Length 42;

Best Local Similarity 59.1%; Pred. No. 1.3;

Matches 13; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

1 WSHPOF-----EKGGSGGGGS 16

10 WRNPOFGGHHHHHGGSGGGGS 31

Search completed: March 2, 2005, 13:02:48
 Job time : 109.098 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 74.6341 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ11
Perfect score: 146
Sequence: 1 wnhpfekgsgsgsggswahpfeK 24

Scoring table: BLOSUM62GX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	80.1	36	14	US-10-026-578B-11
2	71.5	49.0	485	16	US-10-628-432-47
3	71.5	49.0	697	16	US-10-628-432-24
4	69.5	47.6	845	16	US-10-628-432-40
5	69	47.3	24	14	US-10-026-578B-3
6	66	45.2	42	16	US-10-344-607-16
7	63	43.2	36	14	US-10-026-578B-10
8	63	43.2	42	16	US-10-344-607-20
9	61	41.8	646	16	US-10-628-432-49
10	61	41.8	858	16	US-10-628-432-27
11	61	41.8	929	16	US-10-628-432-61
12	59	40.4	117	10	US-09-977-137A-4
13	59	40.4	117	10	US-09-977-137A-5

14	59	40.4	117	10	US-09-977-137A-7	Sequence 7, Appli
15	59	40.4	117	10	US-09-977-137A-8	Sequence 8, Appli
16	59	40.4	117	10	US-09-977-137A-9	Sequence 9, Appli
17	59	40.4	117	10	US-09-977-137A-10	Sequence 10, Appli
18	59	40.4	117	10	US-09-977-137A-11	Sequence 11, Appli
19	59	40.4	118	10	US-09-977-137A-12	Sequence 12, Appli
20	58.5	40.1	81	16	US-10-437-963-197510	Sequence 6, Appli
21	58	39.7	11	15	US-10-354-983-29	Sequence 197510,
22	58	39.7	11	16	US-10-628-432-25	Sequence 29, Appli
23	58	39.7	109	15	US-10-302-100B-41	Sequence 25, Appli
24	58	39.7	661	15	US-10-354-983-8	Sequence 41, Appli
25	58	39.7	763	15	US-10-358-283-15	Sequence 8, Appli
26	57.5	39.4	633	16	US-10-628-432-53	Sequence 15, Appli
27	57	39.0	10	15	US-10-147-211A-20	Sequence 53, Appli
28	56.5	38.7	253	17	US-10-728-246-13	Sequence 20, Appli
29	56	38.4	21	9	US-09-809-517A-30	Sequence 13, Appli
30	56	38.4	22	9	US-09-809-517A-33	Sequence 30, Appli
31	56	38.4	24	9	US-09-809-517A-31	Sequence 33, Appli
32	56	38.4	25	9	US-09-809-517A-34	Sequence 31, Appli
33	56	38.4	129	16	US-10-437-963-161505	Sequence 34, Appli
34	56	38.4	245	17	US-10-887-228A-1	Sequence 161505,
35	56	38.4	246	17	US-10-887-228A-9	Sequence 1, Appli
36	56	38.4	252	17	US-10-887-228A-5	Sequence 9, Appli
37	56	38.4	259	14	US-10-304-630-33	Sequence 5, Appli
38	56	38.4	582	16	US-10-437-963-193053	Sequence 33, Appli
39	56	38.4	795	16	US-10-437-963-108411	Sequence 193053,
40	55.5	38.0	286	16	US-10-437-963-193287	Sequence 108411,
41	55.5	38.0	592	16	US-10-322-281-559	Sequence 193287,
42	55	37.7	204	16	US-10-437-963-161117	Sequence 559, App
43	55	37.7	247	16	US-10-437-963-148362	Sequence 161117,
44	55	37.7	479	16	US-10-437-963-151628	Sequence 148362,
45	55	37.7	959	16	US-10-437-963-181639	Sequence 151628,
						Sequence 181639,

ALIGNMENTS

RESULT 1
US-10-026-578B-11
Sequence 11, Application US/10026578B
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(28)
OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11
Query Match 80.1%; Score 117; DB 14; Length 36;
Best Local Similarity 63.9%; Pred. No. 1.5e-06;


```

; NAME/KEY: MISC FEATURE
; LOCATION: (16)_(16)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-3

Query Match          47.3%; Score 69; DB 14; Length 24;
Best Local Similarity 46.9%; Pred. No. 0.31;
Matches 15; Conservative 0; Mismatches 1; Indels 16; Gaps 2;

QY      1 WSHPOFEKGGSGSGS-----WSHPQFEK 24
Db      1 WSHPOFE-----KXXXXXXXXXXWSHPQFEK 24

RESULT 6
US-10-344-607-16
; Sequence 16, Application US/10344607
; Publication No. US20040161748A1
; GENERAL INFORMATION:
; APPLICANT: HE, MINGYUE
; APPLICANT: TAUSIG, MICHAEL JOHN
; TITLE OF INVENTION: FUNCTIONAL PROTEIN ARRAYS
; FILE REFERENCE: 37945-0049
; CURRENT APPLICATION NUMBER: US/10/344,607
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: PCT/GB01/03657
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-344-607-16

Query Match          45.2%; Score 66; DB 16; Length 42;
Best Local Similarity 59.1%; Pred. No. 1.1;
Matches 13; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY      1 WSHPOF-----EKGGSGSGGS 16
Db      10 WRHPQFGHHHHHGGSGSGGS 31

RESULT 7
US-10-026-578B-10
; Sequence 10, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 10
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
```

```

; NAME/KEY: MISC FEATURE
; LOCATION: (9)_(28)
; OTHER INFORMATION: X represents a single amino acid at each of the positions indicat
; OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are
; OTHER INFORMATION: missing, the total numbers of x will be no less than 5
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10

Query Match          43.2%; Score 63; DB 14; Length 36;
Best Local Similarity 34.1%; Pred. No. 2.2;
Matches 15; Conservative 0; Mismatches 1; Indels 28; Gaps 2;

QY      1 WSHPOFEKGGSGSGS-----WSHPQFEK 24
Db      1 WSHPOFE-----KXXXXXXXXXXXXXXXXXXXXWSHPQFEK 36

RESULT 8
US-10-344-607-20
; Sequence 20, Application US/10344607
; Publication No. US20040161748A1
; GENERAL INFORMATION:
; APPLICANT: HE, MINGYUE
; APPLICANT: TAUSIG, MICHAEL JOHN
; TITLE OF INVENTION: FUNCTIONAL PROTEIN ARRAYS
; FILE REFERENCE: 37945-0049
; CURRENT APPLICATION NUMBER: US/10/344,607
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: PCT/GB01/03657
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-344-607-20

Query Match          43.2%; Score 63; DB 16; Length 42;
Best Local Similarity 54.5%; Pred. No. 2.5;
Matches 12; Conservative 1; Mismatches 1; Indels 8; Gaps 1;

QY      9 GGGSGGG-----SMWHPQF 22
Db      13 GGGSGGGHHHHHSRAWRHPQF 34

RESULT 9
US-10-628-432-49
; Sequence 49, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 49
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct E
US-10-628-432-49
```

Query Match 41.8%; Score 61; DB 16; Length 646;
Best Local Similarity 71.4%; Pred. No. 49;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 11 GSGGSGSMHPQPEK 24
DB 633 GRKGSAMSHPOPEK 646

RESULT 10

US-10-628-432-27
; Sequence 27, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 27
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

Query Match 41.8%; Score 61; DB 16; Length 858;
Best Local Similarity 71.4%; Pred. No. 63;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 11 GSGGSGSMHPQPEK 24
DB 845 GRKGSAMSHPOPEK 858

RESULT 11

US-10-437-963-167129
; Sequence 167129, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barzak, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO: 167129
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_65770C.1.pcp
US-10-437-963-167129

Query Match 41.8%; Score 61; DB 16; Length 929;
Best Local Similarity 52.9%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 WSHPOPEKGGGSGGGSW 17
DB 537 WRSFAMDHGGGSGGGSW 553

RESULT 12

US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Sumner, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match 40.4%; Score 59; DB 10; Length 117;
Best Local Similarity 64.7%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGGSGGSMHPQPEK 24
DB 101 KGNVSCPAMSHPOPEK 117

RESULT 13

US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Sumner, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match 40.4%; Score 59; DB 10; Length 117;
Best Local Similarity 64.7%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGGSGGSMHPQPEK 24
DB 101 KGNVSCPAMSHPOPEK 117

RESULT 14

US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Sumner, Anne O.

```

/ APPLICANT: Caguiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

```

```

Query Match          40.4%; Score 59; DB 10; Length 117;
Best Local Similarity 64.7%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      8 KGGGSGGGSWSHPQFEK 24
      |||:|||||
Db      101 KGNVSCPSAWSHPQFEK 117

```

```

RESULT 15
US-09-977-137A-8
/ Sequence 8, Application US/0977137A
/ Publication No. US20030104524A1
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Caguiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 8
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

```

```

Query Match          40.4%; Score 59; DB 10; Length 117;
Best Local Similarity 64.7%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      8 KGGGSGGGSWSHPQFEK 24
      |||:|||||
Db      101 KGNVSCPSAWSHPQFEK 117

```

Search completed: March 2, 2005, 14:18:52
 Job time : 74.6341 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 27.3171 Seconds
(without alignment)
65.585 Million cell updates/sec

Title: SEQ11
Perfect score: 146
Sequence: 1 wnhpgefkgsgsgsgshpgefek 24

Scoring table:
BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6CTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	40.4	117	4	US-09-977-137A-4
2	59	40.4	117	4	US-09-977-137A-5
3	59	40.4	117	4	US-09-977-137A-7
4	59	40.4	117	4	US-09-977-137A-8
5	59	40.4	117	4	US-09-977-137A-10
6	59	40.4	117	4	US-09-977-137A-12
7	59	40.4	117	4	US-09-977-137A-6
8	59	40.4	118	4	US-09-977-137A-6
9	58	39.7	109	4	US-09-407-687-41
10	56	38.4	21	4	US-09-809-517A-30
11	56	38.4	22	4	US-09-809-517A-33
12	56	38.4	24	4	US-09-809-517A-31
13	56	38.4	25	4	US-09-809-517A-34
14	56	38.4	259	4	US-09-431-887-33
15	56	38.4	660	2	US-08-770-761A-2
16	56	38.4	662	2	US-08-770-761A-5
17	56	38.4	705	2	US-08-770-761A-7
18	54	37.0	117	4	US-09-977-137A-11
19	54	37.0	287	4	US-09-248-796A-15613
20	53	36.3	99	4	US-09-407-687-43
21	52.5	36.0	286	4	US-09-248-796A-15692
22	52	35.6	450	4	US-09-849-016-7527
23	52	35.6	10	4	US-09-809-517A-6
24	52	35.6	521	4	US-09-252-991A-18266
25	51	34.9	563	4	US-09-600-991A-4
26	51	34.9	8	3	US-08-948-097-2
27	51	34.9	8	3	US-09-382-950-7

28	51	34.9	8	3	US-09-382-736B-8	Sequence 8, Appl
29	51	34.9	8	4	US-09-619-103-9	Sequence 9, Appl
30	51	34.9	8	4	US-10-104-218-5	Sequence 5, Appl
31	51	34.9	8	4	US-09-809-517A-9	Sequence 9, Appl
32	51	34.9	26	2	US-08-596-387B-97	Sequence 97, Appl
33	51	34.9	26	3	US-09-067-615-97	Sequence 97, Appl
34	51	34.9	26	5	PCR-US95-09816A-97	Sequence 97, Appl
35	51	34.9	179	4	US-09-270-767-38134	Sequence 38134, A
36	51	34.9	179	4	US-09-270-767-35351	Sequence 5351, A
37	51	34.9	219	4	US-09-809-517A-24	Sequence 27, Appl
38	51	34.9	219	4	US-09-809-517A-27	Sequence 27, Appl
39	51	34.9	229	4	US-09-248-796A-17127	Sequence 17127, A
40	51	34.9	238	4	US-09-495-880A-42	Sequence 42, Appl
41	51	34.9	254	4	US-09-270-767-60227	Sequence 60227, A
42	51	34.9	266	4	US-09-495-880A-26	Sequence 26, Appl
43	51	34.9	328	4	US-09-270-767-44768	Sequence 44768, A
44	51	34.9	699	4	US-09-252-991A-17780	Sequence 17780, A
45	50.5	34.6	256	3	US-09-128-450-22	Sequence 22, Appl

ALIGNMENTS

```

RESULT 1
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match      40.4% Score 59; DB 4; Length 117;
Best Local Similarity 64.7% Pred. No. 3.52;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      8 KGGSGGGSGWSHPQFEK 24
Db      101 KGNVSCPSAWSHPQFEK 117

RESULT 2
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5

```

LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match 40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 3
US-09-977-137A-7
Sequence 7, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Cagliati, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

Query Match 40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 4
US-09-977-137A-8
Sequence 8, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Cagliati, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match 40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 5
US-09-977-137A-9
Sequence 9, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Cagliati, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match 40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 6
US-09-977-137A-10
Sequence 10, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Cagliati, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match 40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24
DB 101 KGNVSCPSAMWHPQFEK 117

```
RESULT 7
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match          40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      8 KGGGSGGSMWHPQFEK 24
Db      101 KGNVSCPSAWSHQFEK 117

RESULT 8
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match          40.4%; Score 59; DB 4; Length 118;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      8 KGGGSGGSMWHPQFEK 24
Db      102 KGNVSCPSAWSHQFEK 118

RESULT 9
US-09-407-687-41
; Sequence 41, Application US/09407687
; Patent No. 6548634
; GENERAL INFORMATION:
```

```
; APPLICANT: Ballinger, Marcus
; APPLICANT: Kavanaugh, Michael
; TITLE OF INVENTION: Synthetic Peptides Having FGF Receptor
; FILE OF INVENTION: Affinity
; FILE REFERENCE: 1517.001
; CURRENT APPLICATION NUMBER: US/09/407,687
; CURRENT FILING DATE: 1999-09-28
; EARLIER APPLICATION NUMBER: 60/102,667
; EARLIER FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 109
; TYPE: PRT
; ORGANISM: SYArtificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-09-407-687-41

Query Match          39.7%; Score 58; DB 4; Length 109;
Best Local Similarity 68.8%; Pred. No. 3.9;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1 WSHPOFEKGGSGGGS 16
        |||: |||||
Db      24 WSHFRNGFGGSGGGS 39

RESULT 10
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
; FILE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match          38.4%; Score 56; DB 4; Length 21;
Best Local Similarity 81.8%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      14 GGSWHPQFEK 24
        |||||
Db      11 GAPWHPQFEK 21

RESULT 11
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
; FILE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
```

```

; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

```

```

Query Match          38.4%; Score 56; DB 4; Length 22;
Best Local Similarity 81.8%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 14 GGSMSHPQPEK 24
DB 12 GAPMSHPQPEK 22

```

```

RESULT 12
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136e1 methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

```

```

Query Match          38.4%; Score 56; DB 4; Length 24;
Best Local Similarity 81.8%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 14 GGSMSHPQPEK 24
DB 14 GAPMSHPQPEK 24

```

```

RESULT 13
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136e1 methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41

```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

```

```

Query Match          38.4%; Score 56; DB 4; Length 25;
Best Local Similarity 81.8%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 14 GGSMSHPQPEK 24
DB 15 GAPMSHPQPEK 25

```

```

RESULT 14
US-09-431-887-33
; Sequence 33, Application US/09431887
; Patent No. 6534036
; GENERAL INFORMATION:
; APPLICANT: D-Gen Limited
; TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
; FILE REFERENCE: ICOT/P21952
; CURRENT APPLICATION NUMBER: US/09/431,887
; PRIOR FILING DATE: 1999-11-02
; PRIOR APPLICATION NUMBER: GB 9824091.4
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Trichosurus vulpecula
US-09-431-887-33

```

```

Query Match          38.4%; Score 56; DB 4; Length 259;
Best Local Similarity 40.7%; Pred. No. 17;
Matches 11; Conservative 2; Mismatches 4; Indels 10; Gaps 2;

```

```

QY 1 MSHPQPEKGGSS-----GGGSMSHP 20
DB 58 MGHPO---GGSTMGQPHPGSNWGP 81

```

```

RESULT 15
US-08-770-761A-2
; Sequence 2, Application US/0870761A
; Patent No. 5814503
; GENERAL INFORMATION:
; APPLICANT: Kovacevic, Steven
; APPLICANT: Otto, Keith A.
; APPLICANT: Rao, Ramachandra N.
; TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE
; TITLE OF INVENTION: REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,761A

```

FILED DATE: 19-DEC-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36, 808
REFERENCE/DOCKET NUMBER: X-10136
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-376-0756
TELEFAX: 317-277-1917
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 660 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-770-761A-2

Query Match 38.4%; Score 56; DB 2; Length 660;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 7 EKGGSGGGSGWSHPQF 22
Db 643 DEGNPEGSSAMRHPQF 658

Search completed: March 2, 2005, 12:25:36
Job time : 27.3171 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 18.3415 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ12
Perfect score: 110
Sequence: 1 wshpgfexxxxxxxxxxwshpgfek 24

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	59	53.6	298	2	D89982	hypothetical prote
2	57	51.8	650	2	T23175	hypothetical prote
3	56.5	51.4	629	2	AF0501	glucose inhibited
4	55	50.0	385	2	D72211	acetylornithine am
5	55	50.0	699	2	A54660	histidine rich cal
6	54	49.1	83	1	C42645	ribosomal protein
7	54	49.1	83	2	F81664	ribosomal protein
8	54	49.1	435	2	A44308	Antho-Rhamide prec
9	53	48.2	147	2	E89838	hypothetical prote
10	53	48.2	268	2	G81305	probable dnaI-like
11	53	48.2	338	2	H96765	hypothetical prote
12	52	47.3	218	2	AH1117	transaldolase homo
13	52	47.3	218	2	AB1478	transaldolase homo
14	52	47.3	397	2	C84904	hypothetical prote
15	52	47.3	471	2	JQ1021	alpha-galactosidas
16	52	47.3	803	2	T46179	hypothetical prote
17	52	47.3	2128	2	I52577	beta-spectrin - mo
18	51	46.4	86	2	C86570	S17 ribosomal prot
19	51	46.4	86	2	H72054	ribosomal protein
20	51	46.4	124	2	A30043	triphoblast-specif
21	51	46.4	430	2	AH3420	phage DNA packagin
22	51	46.4	458	2	T16041	hypothetical prote
23	51	46.4	466	2	AH1126	endo-1,4-beta-xyla
24	51	46.4	481	2	T32991	hypothetical prote
25	51	46.4	537	1	YRHUB6	cytosine-related
26	51	46.4	537	1	YRMSB6	tyrosinase-related
27	51	46.4	547	1	G71307	probable oligopept
28	51	46.4	597	1	HOECT1	hydrogenase (EC 1.
29	51	46.4	597	2	AE0721	hydrogenase-1 larg

30	51	46.4	597	2 B85632	hydrogenase-1 larg
31	51	46.4	597	2 A30770	hydrogenase-1 larg
32	51	46.4	1418	2 T15232	hypothetical prote
33	51	46.4	2469	2 H36812	hypothetical prote
34	50	45.5	103	2 E90973	hypothetical prote
35	50	45.5	103	2 A85821	unknown protein en
36	50	45.5	103	2 C85713	unknown protein en
37	50	45.5	103	2 A30904	hypothetical prote
38	50	45.5	243	2 T08785	hypothetical prote
39	50	45.5	277	2 A97539	MGCI1873 protein ho
40	50	45.5	277	2 AD2758	esterase D (import
41	50	45.5	284	2 AH3479	S-formylglutathion
42	50	45.5	310	2 C84701	hypothetical prote
43	50	45.5	334	2 A39172	Antho-Rhamide neur
44	50	45.5	467	2 C87349	efflux system prot
45	50	45.5	534	2 T30268	oligo-1,6-glucosid

ALIGNMENTS

RESULT 1
D89982 hypothetical protein SA1747 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89982
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsutsu, K.
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
Lancet 357, 1225-1240, 2001
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89982
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <KUR>
A:Cross-references: UNIPROT:Q99SV5; GB:BA000018; PID:G13701726; PIDN:BA843019.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA1747

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 4.1;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXWSHPQFE 23
| : |||:||||| : |||
Db 90 WTKQLEKMIAPFYRKMDHQVE 112

RESULT 2
T23175 hypothetical protein K01C8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T23175
R:Sim, M.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19702
A:Accession: T23175
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-650 <WIL>
A:Cross-references: UNIPROT:Q21087; EMBL:Z49068; PIDN:CAA88862.1; GSPDB:GN00020; CESP:K0
A:Experimental source: clone K01C8
C:Genetics:
A:Gene: CESP:K01C8.3
A:Map position: 2
A:Features: 31/1; 74/1; 181/3; 392/2; 416/1; 489/3; 551/1
F:78-546/Domain: animal histidine decarboxylase homology <HDC>
Query Match 51.8%; Score 57; DB 2; Length 650;

Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of *Chlamydia trachomatis* Moyn and *Chlamydia pneumoniae* AR39.

A:Reference number: AB1500; MUID:20150255; PMID:10684935

A:Accession: F81664

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-83 <TET>

A:Cross-references: UNIPROT:Q9PUM3; GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF3960

A:Experimental source: strain N195 (Mopn)

C:Genetics: TC0806

C:Superfamily: Escherichia coli ribosomal protein S17

Query Match 49.1%; Score 54; DB 2; Length 83;

Best Local Similarity 26.1%; Pred. No. 4.7;

Matches 6; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 23

Db 33 YSHPOYAKVVRDSKYVAINELD 55

RESULT 8

Antho-Ramide precursor - see anemone (*Anthopleura elegantissima*)

C:Species: *Anthopleura elegantissima*

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A44308

R:Schmutzler, C.; Darmer, D.; Diekhoff, D.; Grimmelshuizen, C.J.

J. Biol. Chem. 267, 22534-22541, 1992

A>Title: Identification of a novel type of processing sites in the precursor for the see

A:Reference number: A44308; MUID:93054550; PMID:1429603

A:Accession: A44308

A>Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-435 <SCH>

A:Cross-references: UNIPROT:P10419; GB:M98269; NID:g155702; PIDN:AA27738.1; PID:g155703

A>Note: sequence extracted from NCBI backbone (NCBIN:117102, NCBI:P117104)

C:Keywords: neuropeptide

Query Match 49.1%; Score 54; DB 2; Length 435;

Best Local Similarity 39.3%; Pred. No. 32;

Matches 11; Conservative 9; Mismatches 4; Indels 4; Gaps 1;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 24

Db 66 FSDPQFMKGRFSDPQFMKGRFSDPQFMK 93

RESULT 9

hypothetical protein [imported] - *Staphylococcus aureus* (strain N315)

C:Species: *Staphylococcus aureus*

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C:Accession: E89838

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, M.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:21311952; PMID:11410146

A:Accession: E89838

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-147 <KUR>

A:Cross-references: UNIPROT:Q99VU5; GB:BA000018; PID:g13700567; PIDN:BA81864.1; GSPDB:G

A:Experimental source: strain N315

C:Genetics: SA0631

Query Match 48.2%; Score 53; DB 2; Length 147;

Best Local Similarity 33.3%; Pred. No. 13;

Matches 8; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 24

Db 69 WGHFNSKSMVNIELLYVEPOFRK 92

RESULT 10

probable dnaJ-like protein Cj1034c [imported] - *Campylobacter jejuni* (strain NCTC 11168)

C:Species: *Campylobacter jejuni*

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004

C:Accession: G81305

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell

Nature 403, 665-668, 2000

A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: G81305

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-268 <PAR>

A:Cross-references: UNIPROT:Q9PNQ7; GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CB7325

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics: Cj1034c

Query Match 48.2%; Score 53; DB 2; Length 268;

Best Local Similarity 31.8%; Pred. No. 25;

Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HPQFEKXXXXXXXXXSHPOFE 24

Db 231 HPDFHOGSAIKKAYAREQFEK 252

RESULT 11

hypothetical protein F25P22.26 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: H96765

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.R.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H96765

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <STO>

A:Cross-references: UNIPROT:Q9C9S8; GB:AE005173; NID:g6692747; PIDN:AAF24853.1; GSPDB:G

C:Genetics: F25P22.26

A:Map position: 1

Query Match 48.2%; Score 53; DB 2; Length 338;

Best Local Similarity 38.1%; Pred. No. 33;

Matches 8; Conservative 10; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPO 21

Db 101 FSHPOGHLPAYTSMSPQ 121

RESULT 12

transaldolase homolog lmo0343 [imported] - *Listeria monocytogenes* (strain EGD-e)

C:Species: *Listeria monocytogenes*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
 C:Accession: AH1117
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fahn, H.; Jones, L.M.; Karet, U.
 Science 294, 849-852, 2001
 A:Authors: Krefte, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma, O.K.; Schluteter, T.; Simoes, N.; Tletter, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.
 A:Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AH1117
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-218 <GLA>
 A:Cross-references: UNIPROT:Q92EU7; GB:NC_003210; PIDN:CAC98422.1; PID:916409721; GSPDB:A; Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0343
 C:Superfamily: Bacillus subtilis 23K phosphoprotein orfU

Query Match 47.3%; Score 52; DB 2; Length 218;
 Best Local Similarity 35.3%; Pred. No. 27;
 Matches 6; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXW 17
 DB 193 WSHPLTKGIEGFLKDW 209

RESULT 13
 AB1478
 transaldolase homlog lmo361 [imported] - *Listeria innocua* (strain Clp11262)
 C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
 C:Accession: AB1478
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fahn, H.; Jones, L.M.; Karet, U.
 Science 294, 849-852, 2001
 A:Authors: Krefte, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma, O.K.; Schluteter, T.; Simoes, N.; Tletter, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.
 A:Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AB1478
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-218 <GLA>
 A:Cross-references: UNIPROT:Q92EU7; GB:AL592022; PIDN:CAC95594.1; PID:916412790; GSPDB:A; Experimental source: strain Clp11262
 C:Genetics:
 A:Gene: lmo361
 C:Superfamily: Bacillus subtilis 23K phosphoprotein orfU

Query Match 47.3%; Score 52; DB 2; Length 218;
 Best Local Similarity 35.3%; Pred. No. 27;
 Matches 6; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXW 17
 DB 193 WSHPLTKGIEGFLKDW 209

RESULT 14
 C84904
 hypothetical protein At2g46550 [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
 C:Accession: C84904
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MUID:20083487; PMID:10617157
 A:Accession: C84904
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-397 <STO>
 A:Cross-references: UNIPROT:Q92PR4; GB:AE002093; NID:94415936; PIDN:AAD20166.1; GSPDB:GN
 C:Genetics:
 A:Gene: At2g46550
 A:Map position: 2

Query Match 47.3%; Score 52; DB 2; Length 397;
 Best Local Similarity 34.8%; Pred. No. 54;
 Matches 8; Conservative 10; Mismatches 3; Indels 2; Gaps 1;

QY 3 HPQFEKXXXXXXXXXWSP--QFE 23
 DB 74 NPNFPDLPHNTRLMSHPHOFQ 96

RESULT 15
 JQ1021
 alpha-galactosidase (EC 3.2.1.22) - Yeast (*Saccharomyces cerevisiae*) (strain carlsberg
 C:Species: *Saccharomyces cerevisiae*
 A:Variety: *Saccharomyces carlsbergensis* NCYC396
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 09-Jul-2004
 C:Accession: JQ1021
 R:Turakainen, H.; Korhola, M.; Aho, S.
 Gene 101, 97-104, 1991
 A:Title: Cloning, sequence and chromosomal location of a MEL gene from *Saccharomyces car*
 A:Reference number: JQ1021; MUID:91285441; PMID:1711992
 A:Accession: JQ1021
 A:Molecule type: DNA
 A:Residues: 1-471 <TUR>
 A:Cross-references: UNIPROT:Q03647; GB:M58484; NID:9171923; PIDN:AAA4769.1; PID:9171924
 A:Experimental source: strain NCYC396
 A:Note: the source is designated as *Saccharomyces carlsbergensis*
 C:Comment: Yeast strains producing this enzyme are able to use melibiose as a carbon sou
 C:Genetics:
 A:Gene: MEL
 A:Map position: 10
 A:Superfamily: alpha-galactosidase
 C:Keywords: glycosidase; hydrolase

Query Match 47.3%; Score 52; DB 2; Length 471;
 Best Local Similarity 36.4%; Pred. No. 66;
 Matches 8; Conservative 9; Mismatches 3; Indels 2; Gaps 1;

QY 1 WSH--POFEKXXXXXXXXXWSP 20
 DB 328 WRHQPVDTKYGGGSLQFWSGP 349

Search completed: March 2, 2005, 12:28:58
 Job time : 19.3415 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 89.1707 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ12
Perfect score: 110
Sequence: 1 whpqfexkxxxxxxwhpqfek 24

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	54.5	743	2	08CH90
2	59	53.6	298	2	08NVR8
3	59	53.6	298	2	099SV5
4	59	53.6	298	2	07A4N3
5	59	53.6	298	2	06G800
6	59	53.6	298	2	06GFC4
7	59	53.6	508	1	MATK MARSC
8	59	53.6	508	2	09BBG0
9	59	53.6	508	2	09BBG3
10	59	53.6	704	2	08CUD2
11	57	51.8	452	2	09H7M8
12	57	51.8	504	2	088BY3
13	57	51.8	650	2	021087
14	57	51.8	691	2	08ENR9
15	57	51.8	705	2	095ZS2
16	57	51.8	1714	2	09FMM3
17	56.5	51.4	107	2	06ZJ15
18	56.5	51.4	629	2	GIDA YERPE
19	56.5	51.4	629	2	066JF9
20	55.5	50.5	389	2	09AD61
21	55	50.0	256	2	06GZX4
22	55	50.0	355	2	09T1Z2
23	55	50.0	385	1	ARGD THEME
24	55	50.0	436	2	088TY1
25	55	50.0	699	1	SRCH HUMAN
26	54.5	49.5	1131	2	0881I8
27	54	49.1	83	1	RS17_CHLNU
28	54	49.1	83	1	RS17_CHLTR
29	54	49.1	208	2	062TR3
30	54	49.1	256	2	065IB2
31	54	49.1	265	2	09PWS6

32	54	49.1	429	1	PMR2 ANTEL	Q16994 anthopleura
33	54	49.1	435	1	PMR1 ANTEL	P10419 anthopleura
34	54	49.1	512	2	091402	Q91402 polyandrium
35	54	49.1	535	1	TYR1 CHICK	O57405 gallus gall
36	54	49.1	536	2	091BH8	Q91BH8 coturnix co
37	54	49.1	587	2	08W232	Q9W232 dirosophila
38	54	49.1	587	2	09V9B8	Q9V9B8 dirosophila
39	54	49.1	802	2	07UMP8	Q7UMP8 rhodopirell
40	54	49.1	1676	2	08RG36	Q8RG36 utillago ma
41	53.5	48.6	407	2	06NGV4	Q6NGV4 corynebacte
42	53.5	48.6	848	2	08ZRR6	Q8ZRR6 salmoneila
43	53	48.2	147	2	08NXQ8	Q8NXQ8 staphylococ
44	53	48.2	147	2	093VU5	Q93VU5 staphylococ
45	53	48.2	147	2	07A6X9	Q7A6X9 staphylococ

ALIGNMENTS

RESULT 1					
ID	08CH90	PRELIMINARY;	PRT;	743 AA.	
AC	08CH90;				
DT	01-MAR-2003 (TREMBLrel. 23, Created)				
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	Soluble guanylyl cyclase alpha 2 subunit E219G mutant.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TTSUB=Kidney;				
RA	Nakamura I., Yao Y., Suzuki N.;				
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB097860; BAC44887.1; --				
DR	HSSP; P30803; IAZS.				
DR	GO; GO:0004383; F:guanylate cyclase activity; IEA.				
DR	GO; GO:0016829; F:lyase activity; IEA.				
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.				
DR	InterPro; IPR01054; G:cyclase.				
DR	InterPro; IPR009080; tRNAasn_1a_bind.				
DR	pfam; PF00211; Guanylate_cyc; 1.				
DR	SMART; SM00044; CYCC; 1.				
DR	PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.				
DR	PROSITE; PS0125; GUANYLATE_CYCLASES_2; 1.				
KW	lyase.				
SQ	SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;				
Query Match					
Best Local Similarity 54.5%; Score 60; DB 2; Length 743;					
Matches 8; Conservative 9; Mismatches 0; Indels 0; Gaps 0;					
QY	8 KXXXXXXXXXSHPOFEK 24				
DB	727 ETSLVPRGSHPOFEK 743				
RESULT 2					
ID	08NVR8	PRELIMINARY;	PRT;	298 AA.	
AC	08NVR8;				
DT	01-OCT-2002 (TREMBLrel. 22, Created)				
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	MW1874 protein.				
GN	OrderedLocustNames=MW1874;				
OS	Staphylococcus aureus (strain MW2).				
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
OX	NCBI_Taxid=196620;				
RN	[1]				
RP	SEQUENCE FROM N.A.				

RC STRAIN=MM2;
RA MEDLINE=22040717; PubMed=12044376; DOI=10.1016/S0140-6736(02)08713-5;
RX Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RY Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiratsugu K.,
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.",
RL Lancet 359:1819-1827(2002).
CC -1. SIMILARITY: Belongs to the ABC transporter family.
DR EMBL, AP004828; BAB95739.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity; coupled to transmembrane m. . .; IEA.
DR GO; GO:0003698; F:damaged DNA binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR000432; MutS_C.
DR Pfam; PF00005; ABC_tran; I.
DR ProDom; PD000006; ABC_transporter; 1.
DR ProDom; PD001263; MutS_C; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00693; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34532 MW; DDC7A13478EA7948 CRC64;

Query Match	53.6%	Score 59;	DB 2;	Length 298;
Best Local Similarity	34.8%	Pred. No. 24;		
Matches	8;	Conservative	9;	Mismatches 6;
				Indels 0;
				Gaps 0

```
QY      1 WSHNPQFEKXXXXXXXXXWSHPQFE 23
      10 | : | | | | | | | | | |
Db      90 WTKQLEKMIAPFYRKWDHQVFE 112
```

ID	Q99SV5	PRELIMINARY;	PRT;	298 AA.
AC	Q99SV5;			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Similar to ABC transporter (ATP-binding protein).			
GN	OrderedLocusNames=SAV1933;			
OS	Staphylococcus aureus (strain Mu50 / ATCC 700699).			
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.			
NCBI_Taxid=158878;				
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Mu50 / ATCC 700699;			
RX	MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;			
RA	Kutoda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,			
RA	Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,			
RA	Kanamori M., Macumaru H., Maruyama A., Murakami H., Hasegawa A.,			
RA	Mitsunari-Ii Y., Takahashi N. K., Sawano T., Inoue R.-I., Kaito C.,			
RA	Setimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,			
RA	Kaneshisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,			
RA	Hatchett M., Ogaewara N., Hayaashi H., Hiramatsu K.;			
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus			
RT	aureus.";			
RL	Lancet 357:1225-1240 (2001).			
CC	-1- SIMILARITY: Belongs to the ABC transporter family.			
CC	EMBL; AP003363; BAB58095.1; -.			
DR	PIR; D89982; D89982.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0042656; F:ATPase activity; coupled to transmembrane m. . .; IEA.			
DR	GO; GO:0003664; F:damaged DNA binding; IEA.			
DR	GO; GO:0000166; F:nucleotide binding; IEA.			
DR	GO; GO:0006298; P:mismatch repair; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			

DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR004332; Muts_C.
DR Pfam: PF000005; ABC_tran_1.
DR Prodom: PD000006; ABC_transporter_1.
DR Prodom: PD001263; Muts_C_2.
DR SMART: SM00382; AAA_1.
DR PROSITE: PS00693; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete_proteome.
SQ SEQUENCE 239 AA; 34439 MW; 6FAE51C804BB34 CRC64

Query Match	53.6%	Score 59;	DB 2;	Length 298;
Best Local Similarity	34.8%	Pred. No. 24;		
Matches	8;	Conservative	9;	Mismatches 6;
				Indels 0;
				Gaps 0;

```
QY      1 WSHPQFEKXXXXXXXXXXWSHPQFE 23
        | : | | : : : : | | | |
Db      90 WTKQLEKMIAPFYRKWDHQVFE 112
```

RESULT 4		
Q7A4N3		
ID	Q7A4N3	PRELIMINARY; PRT; 298 AA.
AC	Q7A4N3;	
DT	05-JUL-2004 (TrEMBLrel. 27, Created)	
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)	
DE	SA1747 protein.	
GN	Ordered locus names=SA1747;	
OS	Staphylococcus aureus (strain N315).	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OK	NCBI_taxid=158879;	

RX MEDLINE:13111952; PubMed:11418146; DOI=10.1016/S0140-6736(00)04403-2;
 RA Kuroda M., Ohta T., Uehidiyama I., Baba T., Yuzawa H., Kobayashi I.,
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 R A Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mitunori-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuuchi J.,
 RA Kaneshisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
 RT RT
 "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*."
 RL Lancet 357:1225-1240(2001).
 CC 1- SIMILARITY: Belongs to the ABC transporter family.
 DR ENBL, AP003135; BAB43019.1; "-
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0048226; F:ATPase activity, coupled to transmembrane m...; IEA.
 DR GO; GO:0003684; F:damaged dna binding; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006298; F:ismatch repair; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR000432; MutS_C.
 DR Pfam; PF00005; ABC_tran; I.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR ProDom; PD001263; MutS_C; 2.
 DR SMART; SMO0382; AAA; 1_-
 DR PROSITE; PSS0893; ABC_TRANSPORTER_2; 1.
 DR ATP-binding, Complete proteome.
 KW ATP-binding, Complete proteome.
 QO SSQUENCE 298 AA; 3443 MW; 6FAAE5ICB048B334 CRC64;

Query Match	53.6%	Score 59;	DB 2;	Length 298;
Best Local Similarity	34.8%	Pred. No. 24;		
Matches	8;	Conservative	9;	Mismatches 6;
				Indels 0;
				Gaps 0;

Qy	1	WSHPQPEKXXXXXXXXXXSHPQPE	23
	:	:::	
Pb	90	WTTKLEKMIAPFRKWDQVFE	112

```

RESULT 5
Q6G800 PRELIMINARY; PRT; 298 AA.
ID Q6G800;
AC Q6G800;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE Putative ABC transporter ATP-binding protein.
GN OrderedLocName=SA51857;
OS Staphylococcus aureus (strain MSSA476).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=282459;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15213324; DOI=10.1073/pnas.0402521101;
RA Holten M.T.G., Fell E.J., Lindsey J.A., Peacock S.J., Day N.P.J.,
RA Knight M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,
RA Churcher C., Clark L., Corton C., Cronin A., Doggett J., Dowd L.,
RA Fellwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabinowitch E., Rutherford K.M., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
RA Spratt B.G., Parkhill J.;
RT "Complete genomes of two clinical Staphylococcus aureus strains:
RT evidence for the rapid evolution of virulence and drug resistance.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
EMBL; BX571857; CAG4362.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.
DR GO; GO:0003684; F:damaged DNA binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; I.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34532 MW; 0DC7413478EA7948 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXSHPOFE 23
| : | | | | | | | | | | | | | |
DB 90 WTKQLEKMIAPFYRKMDHVF 112

RESULT 6
Q6GFC4 PRELIMINARY; PRT; 298 AA.
ID Q6GFC4;
AC Q6GFC4;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE Putative ABC transporter ATP-binding protein.
GN OrderedLocName=SA3025;
OS Staphylococcus aureus (strain MRSA252).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=282458;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15213324; DOI=10.1073/pnas.0402521101;
RA Holten M.T.G., Fell E.J., Lindsey J.A., Peacock S.J., Day N.P.J.,

```

```

RA Knight M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,
RA Bason N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Corton C., Cronin A., Doggett J., Dowd L.,
RA Fellwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabinowitch E., Rutherford K.M., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
RA Spratt B.G., Parkhill J.;
RT "Complete genomes of two clinical Staphylococcus aureus strains:
RT evidence for the rapid evolution of virulence and drug resistance.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
EMBL; BX571856; CAG41010.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.
DR GO; GO:0003684; F:damaged DNA binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; I.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34474 MW; 666691085CABCA58 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXSHPOFE 23
| : | | | | | | | | | | | | | |
DB 90 WTKQLEKMIAPFYRKMDHVF 112

RESULT 7
MATK MARSC STANDARD; PRT; 508 AA.
ID MATK MARSC
AC Q8WK9;
DT 28-FEB-2003 (rel. 41, Created)
DT 28-FEB-2003 (rel. 41, Last sequence update)
DE 05-JUL-2004 (rel. 44, Last annotation update)
GN Maturase K (intron maturase).
OS Marathrum schiedeanum.
OC Charophyta.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid I; Malpighiales; Podostemaceae; Marathrum.
OX NCBI_TaxID=116737;
RN [1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Phylogenetic relationships of the aquatic angiosperm family
RT Podostemaceae inferred from matk sequence data.";
RL Submitted (FE8-2000) to the EMBL/GenBank/DDA databases.
CC -1- FUNCTION: Probably assists in splicing chloroplast group II
CC introns (by similarity).
CC -1- SIMILARITY: Belongs to the intron maturase family 2. Matk
CC subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: AB038195; BAB3156.1; -.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
SQ SEQUENCE 508 AA; 61342 MW; ECCF5B416B0ABEC3 CRC64;

Query Match
Best Local Similarity 42.1%; Score 59; DB 1; Length 508;
Matches 8; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

Qy 6 FEKXXXXXXXXXSHPOFEK 24
| : : : : : : : : : : : : : : : :
Db 177 FLRLFFHEYWNWSPHSFEK 195

RESULT 8
Q9BBG0 PRELIMINARY; PRT; 508 AA.
AC Q9BBG0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Maturase K.
GN Name: Maturase K.
OS Vanroyenella plumosa.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside I; Malpighiales; Podostemaceae; Vanroyenella.
OX NCBI_TaxID=51609;
RN [1]_TaxID=51609;
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Intrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
RT Inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048378; BAB3398.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0008380; P:RNA splicing; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam; PF01348; Intron_mature2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast.
SQ SEQUENCE 508 AA; 61288 MW; 90229C45138AF4B6 CRC64;

Query Match
Best Local Similarity 42.1%; Score 59; DB 2; Length 508;
Matches 8; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

Qy 6 FEKXXXXXXXXXSHPOFEK 24
| : : : : : : : : : : : : : : : :
Db 177 FLRLFFHEYWNWSPHSFEK 195

RESULT 9
Q9BBG3 PRELIMINARY; PRT; 508 AA.
AC Q9BBG3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Maturase K.
GN Name: Maturase K.
OS Oseya coultieriana.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

```

```

OC euroside I; Malpighiales; Podostemaceae; Oseya.
OX NCBI_TaxID=51602;
RN [1]_TaxID=51602;
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Intrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
RT Inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048375; BAB3395.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0008380; P:RNA splicing; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam; PF01348; Intron_mature2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast.
SQ SEQUENCE 508 AA; 61400 MW; 45240CD31B54CDF CRC64;

Query Match
Best Local Similarity 42.1%; Score 59; DB 2; Length 508;
Matches 8; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

Qy 6 FEKXXXXXXXXXSHPOFEK 24
| : : : : : : : : : : : : : : : :
Db 177 FLRLFFHEYWNWSPHSFEK 195

RESULT 10
Q8CJD2 PRELIMINARY; PRT; 704 AA.
AC Q8CJD2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Guanylyl cyclase alpha 1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A.
RA Nakamura I., Suzuki N.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB096020; BAC24016.1; -.
DR HSPF; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro: IPR001054; G_cyclase.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW Lyase.
SQ SEQUENCE 704 AA; 79188 MW; F73FC97B6854E6A CRC64;

Query Match
Best Local Similarity 50.0%; Score 59; DB 2; Length 704;
Matches 8; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 9 XXXXXXXXXXXXSHPOFEK 24
| : : : : : : : : : : : : : : : :
Db 689 VDLVPRGSMWSPHSFEK 704

RESULT 11
Q9H7M8 PRELIMINARY; PRT; 452 AA.
AC Q9H7M8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

```

DE FLJ00042 protein (Fragment).
GN Name=FLJ00042;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Splice;
RA PubMed:14702039; DOI=10.1038/ng1285;
RA Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Magatsuma M., Shiraori A.,
RA Sudo H., Hoshino T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hoshino T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togaya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshioka Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Montiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hara H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
DR EMBL: AK024450; BAB15740.1; -.
DR GO: GO:0005525; P:small GTPase binding; IEA.
DR GO: GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro: IPR001906; RAS_craftering.
DR PRINTS: PR00449; RASTRNSFRMNG.
FT NON_TER
SQ SEQUENCE 452 AA; 48320 MW; FD10D8A4EA1652B7 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 452;
Best Local Similarity 34.8%; Pred. No. 72;
Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 1 MSHPOPEKXXXXXXXXXMSHPOFE 23
Db 77 MSHPALSCDPTSPERWGVQSE 99

RESULT 12
O88PY3 PRELIMINARY; PRT; 504 AA.
AC O88PY3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Outer membrane efflux protein.
GN OrderedLocusNames=PP0715;
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OC NCBI_TaxID=160486;
OX [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weibel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L.M., Beaman M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA Madupu R., Nelson M.C., White O., Peterson J.D., Khouri H.M.,
RA Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
RA Morazzz A., Utechtack T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA Medler H., Lauber J., Stjepandic D., Hohlsel J., Streitz M., Helm S.,
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Frazer C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL: AE016776; AA065340.1; -.
DR TIGR: PP0715; -.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003423; OSP.
DR InterPro: IPR010331; RND_outter_NodT.
DR Pfam: PF02321; OSP; 2. RND_outter_NodT.
DR TIGRPFAMs: TIGR01845; RND_outter_NodT; 1.
KW Complete proteome.
SQ SEQUENCE 504 AA; 55514 MW; 581092F1804FC14D CRC64;

Query Match 51.8%; Score 57; DB 2; Length 504;
Best Local Similarity 38.1%; Pred. No. 81;
Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

OY 4 POFKXXXXXXXXXMSHPOFE 24
Db 22 POFKQAPWLDGWTPTMIEQ 42

RESULT 13
O21087 PRELIMINARY; PRT; 650 AA.
ID O21087
AC O21087;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein K01C8.3a.
GN Name=tdc-1; Synonyms=K01C8.3a; ORFNames=K01C8.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Felodertinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RA Sims M.A.;
RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
CC -I- COFACTOR: Pyridoxal phosphate (By similarity).
CC -I- SIMILARITY: Belongs to the group II decarboxylase family.
DR EMBL: Z49068; CA88862.1; -.
DR FIR: T2175; T23175.
DR HSSP: PB0041; JUS3.
DR WormBase: WBGene0000562; tdc-1.
DR WormPep: K01C8.3a; CE21011.
DR GO: GO:0016831; P:carboxy-lyase activity; IEA.
DR GO: GO:0016825; P:lyase activity; IEA.
DR GO: GO:0006520; P:amino acid metabolism; IEA.
DR InterPro: IPR010977; Pyridoxal_dec.
DR InterPro: IPR002129; Pyridoxal_dec.
DR Pfam: PF00282; Pyridoxal_dec; 1.
DR PRINTS: PR00800; YHDCRBOXLASE.

```

DR PROSITE; PS00392; DDC GAD HDC YDC; 1.
 KW Hypothetical protein; Lyase; Pyridoxal phosphate.
 SQ SEQUENCE 650 AA; 73201 MW; 4830ECCAPD1AFAB1 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 650;
 Best Local Similarity 41.2%; Pred. No. 1.1e+02;
 Matches 7; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 6 FEKXXXXXXXXXSHPOF 22
 DB 136 FEKLIMPGITMCHPERF 152

RESULT 14

O8ENA9 PRELIMINARY; PRT; 691 AA.

AC O8ENA9; (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical conserved protein.
 GN OrderedLocustNames=OB2578;
 OS Oceanobacillus thelyensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 CX NCBI_TaxID=182710;

QY 11 SEQUENCE FROM N.A.

RC STRAIN=HTE831;
 RX MEDLINE=22220767; PubMed=12235376; DOI=10.1093/nar/gkf526;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 DR EMBL; AP004601; BAC14534.1.
 DR InterPro; IPR004879; DUF235.
 DR InterPro; IPR008928; Glyco_trans_6hp.
 DR Pfam; PF03190; DUF255; 1.
 DR Complete proteome; Hypothetical protein.
 KW KMW COMPLETE PROTEOME; 11EP02E245B06B5 CRC64;
 SQ SEQUENCE 691 AA; 79588 MW; 11EP02E245B06B5 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 691;
 Best Local Similarity 33.3%; Pred. No. 1.2e+02;
 Matches 8; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXSHPOF 24
 DB 247 WDVGCGVRYATDRKXMPHPHEK 270

RESULT 15

O95ZS2 PRELIMINARY; PRT; 705 AA.

AC O95ZS2;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE Hypothetical protein K01C8.3b;
 GN Name=cdc-1; Synonyms=K01C8.3b; ORFNames=K01C8.3;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderae; Caenorhabditis.
 CX NCBI_TaxID=6239;

QY 11 SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;
 RA Sims M.A.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CORACOR: Pyridoxal phosphate (By similarity).
 CC -1- SIMILARITY: Belongs to the group II decarboxylase family.

DR EMBL; Z49068; CAC42319.1; --
 DR PIR; T23168; T23168.
 DR HSSP; P80041; JUS3.
 DR WormBase; WBGene00006562; cdc-1.
 DR GO; GO:0016831; F:carboxy-lyase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0005520; P:amino acid metabolism; IEA.
 DR InterPro; IPR010977; Aromatic_dec.
 DR InterPro; IPR002129; Pyridoxal_dec.
 DR Pfam; PF0282; Pyridoxal_dec; 1.
 DR PRINTS; PR00800; YHDCRBOXLASE.
 DR PROSITE; PS00392; DDC GAD HDC YDC; 1.
 KW Hypothetical protein; Lyase; Pyridoxal phosphate.
 SQ SEQUENCE 705 AA; 79710 MW; BBBBDP83582FE6 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 705;
 Best Local Similarity 41.2%; Pred. No. 1.2e+02;
 Matches 7; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 6 FEKXXXXXXXXXSHPOF 22
 DB 136 FEKLIMPGITMCHPERF 152

Search completed: March 2, 2005, 12:44:26
 Job time : 91.1707 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 108.098 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ12
Perfect score: 110
Sequence: 1 whnpqfekxxxxxxxxwhnpqfek 24

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp19808:*\n2: geneseqp19808:*\n3: geneseqp20008:*\n4: geneseqp20018:*\n5: geneseqp20028:*\n6: geneseqp20038:*\n7: geneseqp20038:*\n8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	24	ABP60362	Abp60362 Streptavi
2	94.5	85.9	35	ABP60369	Abp60369 Streptavi
3	94	85.5	36	ABP60370	Abp60370 Streptavi
4	58.2	117	5	AAU97558	AAU97558 Synthetic
5	58.2	117	5	AAU97553	AAU97553 Synthetic
6	58.2	117	5	AAU97557	AAU97557 Synthetic
7	58.2	117	5	AAU97559	AAU97559 Synthetic
8	58.2	117	5	AAU97555	AAU97555 Synthetic
9	58.2	117	5	AAU97556	AAU97556 Synthetic
10	58.2	117	5	AAU97552	AAU97552 Synthetic
11	58.2	117	5	AAU97554	AAU97554 Synthetic
12	57.3	118	5	AAU97554	AAU97554 Synthetic
13	57.3	118	5	AAU97554	AAU97554 Synthetic
14	57.3	254	2	AAW93967	AAW93967 Plasmid p
15	57.3	254	2	AAW93969	AAW93969 Plasmid p
16	57.3	396	2	AAW93966	AAW93966 Plasmid p
17	57.3	396	4	AAW93966	AAW93966 Plasmid p
18	57.3	396	4	AAW93966	AAW93966 Plasmid p
19	57.3	448	4	AAW93966	AAW93966 Plasmid p
20	57.3	448	4	AAW93966	AAW93966 Plasmid p
21	57.3	448	4	AAW93966	AAW93966 Plasmid p
22	57.3	448	4	AAW93966	AAW93966 Plasmid p
23	57.3	448	4	AAW93966	AAW93966 Plasmid p
24	57.3	448	4	AAW93966	AAW93966 Plasmid p
25	57.3	448	4	AAW93966	AAW93966 Plasmid p

26	61.5	55.9	697	8	ADS20227	Ads20227 Human agg
27	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
28	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
29	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
30	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
31	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
32	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
33	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
34	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
35	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
36	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
37	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
38	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
39	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
40	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
41	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
42	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
43	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
44	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
45	61	55.5	183	6	ADA27294	Ada27294 Plasmid p

ALIGNMENTS

RESULT 1	ABP60362	ABP60362 standard; peptide; 24 AA.
XX	XX	XX
AC	ABP60362;	
XX	XX	XX
DT	28-MAR-2003 (first entry)	
XX	XX	XX
DE	Streptavidin binding peptide SEQ ID NO 3.	
XX	XX	XX
KW	Streptavidin; protein chip; microtitre plate; detection.	
XX	XX	XX
OS	Synthetic.	
XX	XX	XX
FH	Key	Location/Qualifiers
FT	Misc-difference 9..16	
FT	/label= unknown	
XX	XX	XX
PN	DE10113776-A1.	
XX	XX	XX
PD	02-OCT-2002.	
XX	XX	XX
PF	21-MAR-2001; 2001DE-01013776.	
XX	XX	XX
PR	21-MAR-2001; 2001DE-01013776.	
XX	XX	XX
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX	XX	XX
PI	Schmidt T;	
XX	XX	XX
DR	WPI, 2003-031166/03.	
XX	XX	XX
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
PT	modules.	
XX	XX	XX
PS	Disclosure; Page 4; 18pp; German.	
XX	XX	XX
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily	

CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
SQ Sequence 24 AA;

Query Match 100.0%; Score 110; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXWSHPQFEK 24
DB 1 WSHPOFEKXXXXXXXXXWSHPQFEK 24

RESULT 2
ABP60369
ID ABP60369 standard; peptide; 35 AA.
XX
AC ABP60369;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 10.
XX
KM Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.

XX
XX Key Location/Qualifiers
FH MISC-difference 9..27
FT /label= unknown
FT /note= "optionally deleted for 1-15 residues"
XX
XX DE10113776-A1.

XX
XX 02-OCT-2002..
XX
XX 21-MAR-2001; 2001DE-01013776.
XX
XX 21-MAR-2001; 2001DE-01013776.
XX

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
XX WPI; 2003-031166/03.

XX
XX New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.

PS Claim 7; Page 16; 18pp; German.

XX
XX The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
SQ Sequence 35 AA;

Query Match 85.9%; Score 94.5; DB 6; Length 35;

Best Local Similarity 68.6%; Pred. No. 4.4e-05;
Matches 24; Conservative 0; Mismatches 0; Indels 11; Gaps 1;

OY 1 WSHPOFEK-----XXXXXXXXXWSHPQFEK 24
DB 1 WSHPOFEKXXXXXXXXXWSHPQFEK 35

RESULT 3
ABP60370
ID ABP60370 standard; peptide; 36 AA.
XX
AC ABP60370;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 11.
XX
KM Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.

XX
XX Key Location/Qualifiers
FH Region 9..28
FT /note= "GGGS repeats 2-5 optionally absent, residues 13-
FT 28"
FT 9..12
FT /label= GGGS_repeat

XX
XX DE10113776-A1.
XX
XX 02-OCT-2002..
XX
XX 21-MAR-2001; 2001DE-01013776.
XX
XX 21-MAR-2001; 2001DE-01013776.
XX

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
XX WPI; 2003-031166/03.

XX
XX New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.

PS Claim 8; Page 16; 18pp; German.

XX
XX The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
SQ Sequence 36 AA;

Query Match 85.5%; Score 94; DB 6; Length 36;
Best Local Similarity 44.4%; Pred. No. 5.3e-05;
Matches 16; Conservative 8; Mismatches 0; Indels 12; Gaps 1;

OY 1 WSHPOFEK-----XXXXXXXXXWSHPQFEK 24
DB 1 WSHPOFEKGGSGGGSGGGSGGGSGGGSGWSHPQFEK 36

Query Match	Sequence 117 AA;	58.2%	Score 64;	DB 5;	Length 117;
Best Local Similarity	52.9%;	Fred. No. 3;			
Matches 9;	Conservative 8;	Mismatches 0;	Indels 0;	Gaps 0	

```
QY      8 XXXXXXXXXXXXSHPOPEK 24  
        |:::.....|  
Db      -101 KGNVSCPSAMSHPOPEK 117  
  
RESULT 5  
AAU97553  
ID      AAU97553 standard; protein; 117 AA.  
AC      AAU97553;  
XX      XX  
DT      13-AUG-2002 (first entry)  
XX      XX  
DB      Synthetic cadmium/mercury ion binding chelon protein #1.  
KM      Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW      irrigation water; waste stream; contaminated aqueous medium;  
KW      biological fluid; gastrointestinal tract; chelon protein;  
KM      enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW      heavy metal binding protein.  
XX      XX  
OS      Synthetic.  
PN      WO200230962-A2.  
PD      18-APR-2002.  
PF      12-OCT-2001; 2001WO-US031819.  
PR      12-OCT-2000; 2000US-0240465P.  
PA      (UYGE-) UNIV GEORGIA RES FOUND INC.  
PI      Summers AO, Caguiat JT;  
XX      WPI, 2002-435437/46.  
DR      WPI, 2002-435437/46.  
PT      Novel non-naturally occurring recombinant DNA molecule encoding a chelon  
PR      protein useful for binding divalent cation mercury from contaminated  
PT      soil, water, aqueous medium including biological fluids.  
XX      XX  
PS      Claim 4; Page 22; 42pp; English.  
XX      XX  
CC      The present invention relates to a new non-naturally occurring  
CC      recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC      which binds mercuric ions. The invention is useful for recombinantly  
CC      producing a protein in a host-cell, by infecting or transforming a host  
CC      cell capable of expressing a chelon coding sequence with a vector  
CC      comprising a promoter active in the host cell operably linked to a coding  
CC      region for the protein to produce a recombinant host cell and culturing  
CC      the recombinant host cell under conditions, where DNA is expressed. The  
CC      nucleic acid encoding the chelon protein is useful for binding divalent  
CC      mercuric ions, to take up, sequester and concentrate the heavy metal ions  
CC      from contaminated soil, ground water, hydroponic solutions or irrigation  
CC      water of waste streams. The DNA of the invention, when immobilised onto a  
CC      solid support, is useful for concentrating heavy metal ions from  
CC      contaminated environment waste streams or contaminated aqueous medium  
CC      including biological fluids. The nucleic acid, when recombinantly  
CC      expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic),  
CC      is suitable for use in the in vivo sequestration and elimination of  
CC      mercuric ion from gastrointestinal tracts of animals or humans exposed to  
CC      toxic metal ions such as mercury and/or cadmium. The molecules of the  
CC      invention are also useful in water treatment resins. The nucleic acid of  
CC      the invention is highly specific and binds divalent cation such as  
CC      mercury or cadmium with high affinity. The present amino acid sequence  
CC      represents one of a collection (AAU97553-AAU97560) of synthetic  
CC      cadmium/mercury ion binding chelon proteins of the invention. This  
CC      sequence is one of the heavy metal binding proteins termed chelons of the  
CC      invention  
XX      XX  
SQ      Sequence 117 AA;
```


CC sequence is one of the heavy metal binding proteins termed chelons of the
 CC invention

XX Sequence 117 AA;

Query Match 58.2%; Score 64; DB 5; Length 117;

Best Local Similarity 52.9%; Pred. No. 3;

Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOFEK 24

Db 101 KGNVCPSPSAMSHPOFEK 117

RESULT 8

AAU97555

ID AAU97555 standard; protein; 117 AA.

XX AAU97555;

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #3.

KM Mercuric ion; contaminated soil; ground water; hydroponic solution;

KM irrigation water; waste stream; contaminated aqueous medium;

KM biological fluid; gastrointestinal tract; chelon protein;

KM enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;

KM heavy metal binding protein.

OS Synthetic.

PN WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819.

PR 12-OCT-2000; 2000US-0240465P.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Sumners AO, Caguiat JJ;

DR WPI; 2002-435437/46.

PT Novel non-naturally occurring recombinant DNA molecule encoding a chelon
 PT protein useful for binding divalent cation mercury from contaminated
 PT soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein
 CC which binds mercuric ions. The invention is useful for recombinantly
 CC producing a protein in a host-cell, by infecting or transforming a host
 CC cell capable of expressing a chelon coding sequence with a vector
 CC comprising a promoter active in the host cell operably linked to a coding
 CC region for the protein to produce a recombinant host cell and culturing
 CC the recombinant host cell under conditions, where DNA is expressed. The
 CC nucleic acid encoding the chelon protein is useful for binding divalent
 CC mercuric ions, to take up, sequester and concentrate the heavy metal ions
 CC from contaminated soil, ground water, hydroponic solutions or irrigation
 CC water of waste streams. The DNA of the invention, when immobilised onto a
 CC solid support, is useful for concentrating heavy metal ions from
 CC contaminated environment waste streams or contaminated aqueous medium
 CC including biological fluids. The nucleic acid, when recombinantly
 CC expressed in enteric bacteria (which are nonpathogenic and nonpathogenic),
 CC is suitable for use in the in vivo sequestration and elimination of
 CC mercuric ion from gastrointestinal tracts of animals or humans exposed to
 CC toxic metal ions such as mercury and/or cadmium. The molecules of the
 CC invention are also useful in water treatment resins. The nucleic acid of
 CC the invention is highly specific and binds divalent cation such as

CC mercury or cadmium with high affinity. The present amino acid sequence
 CC represents one of a collection (AAU97553-AAU97560) of synthetic
 CC cadmium/mercury ion binding chelon proteins of the invention. This
 CC sequence is one of the heavy metal binding proteins termed chelons of the
 CC invention

XX Sequence 117 AA;

Query Match 58.2%; Score 64; DB 5; Length 117;

Best Local Similarity 52.9%; Pred. No. 3;

Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOFEK 24

Db 101 KGNVCPSPSAMSHPOFEK 117

RESULT 9

AAU97556

ID AAU97556 standard; protein; 117 AA.

XX AAU97556;

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #4.

KM Mercuric ion; contaminated soil; ground water; hydroponic solution;

KM irrigation water; waste stream; contaminated aqueous medium;

KM biological fluid; gastrointestinal tract; chelon protein;

KM enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;

KM heavy metal binding protein.

OS Synthetic.

PN WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819.

PR 12-OCT-2000; 2000US-0240465P.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Sumners AO, Caguiat JJ;

DR WPI; 2002-435437/46.

PT Novel non-naturally occurring recombinant DNA molecule encoding a chelon
 PT protein useful for binding divalent cation mercury from contaminated
 PT soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein
 CC which binds mercuric ions. The invention is useful for recombinantly
 CC producing a protein in a host-cell, by infecting or transforming a host
 CC cell capable of expressing a chelon coding sequence with a vector
 CC comprising a promoter active in the host cell operably linked to a coding
 CC region for the protein to produce a recombinant host cell and culturing
 CC the recombinant host cell under conditions, where DNA is expressed. The
 CC nucleic acid encoding the chelon protein is useful for binding divalent
 CC mercuric ions, to take up, sequester and concentrate the heavy metal ions
 CC from contaminated soil, ground water, hydroponic solutions or irrigation
 CC water of waste streams. The DNA of the invention, when immobilised onto a
 CC solid support, is useful for concentrating heavy metal ions from
 CC contaminated environment waste streams or contaminated aqueous medium
 CC including biological fluids. The nucleic acid, when recombinantly
 CC expressed in enteric bacteria (which are nonpathogenic and nonpathogenic),
 CC is suitable for use in the in vivo sequestration and elimination of
 CC mercuric ion from gastrointestinal tracts of animals or humans exposed to

CC water of waste streams. The DNA of the invention, when immobilised onto a
 CC solid support, is useful for concentrating heavy metal ions from
 CC contaminated environment waste streams or contaminated aqueous medium
 CC including biological fluids. The nucleic acid, when recombinantly
 CC expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic),
 CC is suitable for use in the in vivo sequestration and elimination of
 CC mercuric ion from gastrointestinal tracts of animals or humans exposed to
 CC toxic metal ions such as mercury and/or cadmium. The molecules of the
 CC invention are also useful in water treatment resins. The nucleic acid of
 CC the invention is highly specific and binds divalent cation such as
 CC mercury or cadmium with high affinity. The present amino acid sequence
 CC represents one of a collection (AAU97553-AAU97560) of synthetic
 CC cadmium/mercury ion binding chelon proteins of the invention. This
 CC sequence is one of the heavy metal binding proteins termed chelons of the
 CC invention

CC SQ Sequence 117 AA;

Query Match 58.2%; Score 64; DB 5; Length 117;

Best Local Similarity 52.9%; Pred. No. 3;
 Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOPEK 24

DB 101 KGNVSCPSAMSHPOPEK 117

RESULT 12

AAU97554

ID AAU97554 standard; protein; 118 AA.

AC AAU97554;

DT 13-AUG-2002; (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #2.

KM Mercuric ion; contaminated soil; ground water; hydroponic solution;
 KM irrigation water; waste stream; contaminated aqueous medium;
 KM biological fluid; gastrointestinal tract; chelon protein;

KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;
 KM heavy metal binding protein.

KM Synthetic.

OS WO200230962-A2.

FN 18-APR-2002.

PD 12-OCT-2001; 2001WO-US031819.

PF 12-OCT-2001; 2000US-024046SP.

PR 12-OCT-2000; 2000US-024046SP.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Summers AO, Caguiat JF;

XX WPI; 2002-435437/46.

DR Novel non-naturally occurring recombinant DNA molecule encoding a chelon
 PT protein useful for binding divalent cation mercury from contaminated
 PT soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

PS The present invention relates to a new non-naturally occurring
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein
 CC which binds mercuric ions. The invention is useful for recombinantly
 CC producing a protein in a host-cell, by infecting or transforming a host
 CC cell capable of expressing a chelon coding sequence with a vector
 CC comprising a promoter active in the host cell operably linked to a coding
 CC region for the protein to produce a recombinant host cell and culturing
 CC the recombinant host cell under conditions, where DNA is expressed. The

CC nucleic acid encoding the chelon protein is useful for binding divalent
 CC mercuric ions, to take up, sequester and concentrate the heavy metal ions
 CC from contaminated soil, ground water, hydroponic solutions or irrigation
 CC water of waste streams. The DNA of the invention, when immobilised onto a
 CC solid support, is useful for concentrating heavy metal ions from
 CC contaminated environment waste streams or contaminated aqueous medium
 CC including biological fluids. The nucleic acid, when recombinantly
 CC expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic),
 CC is suitable for use in the in vivo sequestration and elimination of
 CC mercuric ion from gastrointestinal tracts of animals or humans exposed to
 CC toxic metal ions such as mercury and/or cadmium. The molecules of the
 CC invention are also useful in water treatment resins. The nucleic acid of
 CC the invention is highly specific and binds divalent cation such as
 CC mercury or cadmium with high affinity. The present amino acid sequence
 CC represents one of a collection (AAU97553-AAU97560) of synthetic
 CC cadmium/mercury ion binding chelon proteins of the invention. This
 CC sequence is one of the heavy metal binding proteins termed chelons of the
 CC invention

CC SQ Sequence 118 AA;

Query Match 58.2%; Score 64; DB 5; Length 118;

Best Local Similarity 52.9%; Pred. No. 3;
 Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOPEK 24

DB 102 KGNVSCPSAMSHPOPEK 118

RESULT 13

AAW93967

ID AAW93967 standard; protein; 205 AA.

AC AAW93967;

DT 02-JUL-1999 (first entry)

DE Plasmid pBBP21 protein fragment.

KM Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;
 KM tissue-specific surface molecule; tumour-specific surface molecule;

KW tumour imaging; bilin binding protein; Dbc protein.

KW Synthetic.

OS Key Location/Qualifiers

FN Peptide 1..21

FT Protein 22..205

FT /label= fusion_peptide

FT /note= "This fusion peptide is constructed from bilin
 binding protein and a strep-Tag II region."

PN MO9916873-A1.

PD 08-APR-1999.

PF 25-SEP-1998; 98WO-DE002898.

PR 26-SEP-1997; 97DE-01042706.

PA (SKER/) SKERRA A.

PI Skerra A, Beeste G, Schmidt F, Stibora T;

XX WPI; 1999-255093/21.

DR N-PSDB; AAX24100.

PS New anticalline produced by mutation of lipocalin family protein.
 PS Disclosure; Page 70-72; 80pp; German.

CC molecular weight compounds, particularly in assays where Dig is being
CC used as a label. Compared with Dig-specific antibodies, (1) have a
CC simpler structure and are easier to prepare. They have very high
CC specificity for Dig, relative to other steroids, and fusion partners may
CC be attached to either end without compromising their ability to bind
CC ligand

XX
SQ Sequence 254 AA;

Query Match

Best Local Similarity 57.3%; Score 63; DB 4; Length 254;

Matches 9; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

Qy 7 EKXXXXXXXXXMSHPOPEK 24

Db 188 EAACKVNNNSMSHPOPEK 205

Search completed: March 2, 2005, 13:02:48
Job time : 108.098 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 74.6341 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ12
Perfect score: 110
Sequence: 1 whpghfexkxxxxxxwhpghfex 24

Scoring table: BLOSUM62GX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2:	/cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3:	/cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4:	/cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5:	/cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6:	/cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7:	/cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8:	/cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9:	/cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10:	/cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep:*
11:	/cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12:	/cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13:	/cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep:*
14:	/cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15:	/cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16:	/cgn2_6/ptodata/2/pubppaa/US10D_PUBCOMB.pep:*
17:	/cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
18:	/cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
19:	/cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
20:	/cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	24	US-10-026-578B-3	Sequence 3, Appli
2	94	85.5	36	US-10-026-578B-10	Sequence 10, Appli
3	81	73.6	36	US-10-026-578B-11	Sequence 11, Appli
4	81	58.2	21	US-09-809-517A-30	Sequence 30, Appli
5	64	58.2	22	US-09-809-517A-33	Sequence 33, Appli
6	64	58.2	117	US-09-977-137A-4	Sequence 4, Appli
7	64	58.2	117	US-09-977-137A-5	Sequence 5, Appli
8	64	58.2	117	US-09-977-137A-7	Sequence 7, Appli
9	64	58.2	117	US-09-977-137A-8	Sequence 8, Appli
10	64	58.2	117	US-09-977-137A-9	Sequence 9, Appli
11	64	58.2	117	US-09-977-137A-10	Sequence 10, Appli
12	64	58.2	117	US-09-977-137A-11	Sequence 11, Appli
13	64	58.2	117	US-09-977-137A-12	Sequence 12, Appli

14	64	58.2	118	10	US-09-977-137A-6	Sequence 6, Appli
15	62	56.4	845	16	US-10-628-432-40	Sequence 40, Appli
16	61.5	55.9	485	16	US-10-628-432-47	Sequence 47, Appli
17	61.5	55.9	697	16	US-10-628-432-24	Sequence 24, Appli
18	61	55.5	24	9	US-09-809-517A-31	Sequence 31, Appli
19	61	55.5	25	9	US-09-809-517A-34	Sequence 34, Appli
20	61	55.5	245	17	US-10-887-228A-1	Sequence 1, Appli
21	61	55.5	246	17	US-10-887-228A-9	Sequence 9, Appli
22	61	55.5	252	17	US-10-887-228A-5	Sequence 5, Appli
23	60	54.5	646	16	US-10-628-432-49	Sequence 49, Appli
24	60	54.5	661	15	US-10-354-983-8	Sequence 8, Appli
25	60	54.5	763	15	US-10-358-283-15	Sequence 15, Appli
26	60	54.5	858	16	US-10-628-432-27	Sequence 27, Appli
27	59.5	54.1	633	16	US-10-628-432-53	Sequence 53, Appli
28	59	53.6	24	14	US-10-026-578B-4	Sequence 4, Appli
29	59	53.6	126	9	US-09-815-242-12508	Sequence 12508, A
30	59	53.6	231	9	US-09-815-242-5760	Sequence 5760, Ap
31	59	53.6	298	15	US-10-282-122A-44219	Sequence 44219, A
32	57	51.8	650	15	US-10-369-493-5480	Sequence 5480, Ap
33	57	51.8	847	15	US-10-389-566-832	Sequence 832, App
34	57	51.8	1059	15	US-10-389-566-834	Sequence 834, App
35	57	51.8	1069	15	US-10-389-566-833	Sequence 833, App
36	57	51.8	1069	16	US-10-437-963-143424	Sequence 143424, A
37	56.5	51.4	134	16	US-10-437-963-104065	Sequence 104065, A
38	56.5	51.4	252	16	US-10-437-963-200150	Sequence 200150, A
39	56.5	51.4	629	15	US-10-282-122A-77703	Sequence 77703, A
40	55.5	50.5	158	15	US-10-424-599-184769	Sequence 184769, A
41	55.5	50.5	162	14	US-10-204-887-94	Sequence 94, Appli
42	55.5	50.5	212	15	US-10-424-559-184075	Sequence 184075, A
43	55	50.0	56	11	US-09-833-245-1103	Sequence 1103, Ap
44	55	50.0	56	11	US-09-833-245-1105	Sequence 1105, Ap
45	55	50.0	329	15	US-10-449-831A-204	Sequence 204, App

ALIGNMENTS

RESULT 1
US-10-026-578B-3
Sequence 3, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (10)..(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:

```
NAME/KEY: MISC FEATURE
LOCATION: (12)..(12)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (13)..(13)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (14)..(14)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (15)..(15)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (16)..(16)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-3
```

Query Match 100.0%; Score 110; DB 14; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 24
1 WSHPOFEKXXXXXXXXXXWSHPQFEK 24
```

```
RESULT 2
US-10-026-578B-10
Sequence 10, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC FEATURE
LOCATION: (9)..(28)
OTHER INFORMATION: X represents a single amino acid at each of the positions indicat
OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are
OTHER INFORMATION: missing, the total numbers of x will be no less than 5
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10
```

Query Match 85.5%; Score 94; DB 14; Length 36;
Best Local Similarity 66.7%; Pred. No. 0.00012;
Matches 24; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

```
OY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 24
1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 24
```

```
DB 1 WSHPOFEKXXXXXXXXXXXXXXXXXXWSHPQFEK 36
```

```
RESULT 3
US-10-026-578B-11
Sequence 11, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC FEATURE
LOCATION: (9)..(28)
OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11
```

Query Match 73.6%; Score 81; DB 14; Length 36;
Best Local Similarity 41.7%; Pred. No. 0.0058;
Matches 15; Conservative 8; Mismatches 1; Indels 12; Gaps 1;

```
OY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 24
1 WSHPOFEKGGSGGGSGGGSGGGSGGGSGSHPOFEK 36
```

```
RESULT 4
US-09-809-517A-30
Sequence 30, Application US/09809517A
Patent No. US20020034733A1
GENERAL INFORMATION:
APPLICANT: lohnig, Corinna
TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 30
LENGTH: 21
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
```

Query Match 58.2%; Score 64; DB 9; Length 21;
Best Local Similarity 52.9%; Pred. No. 0.49;


```

; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVSCPSAMSHPOFEK 117

RESULT 10
US-09-977-137A-9
; Sequence 9; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVSCPSAMSHPOFEK 117

RESULT 11
US-09-977-137A-10
; Sequence 10; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10
```

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVSCPSAMSHPOFEK 117

RESULT 12
US-09-977-137A-11
; Sequence 11; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVPCPSAMSHPOFEK 117

RESULT 13
US-09-977-137A-12
; Sequence 12; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12
```

Search completed: March 2, 2005, 14:18:59
 Job time : 81.6341 secs

FEATURE: 1
 OTHER INFORMATION: Description of Artificial Sequence: chelon
 US-09-977-137A-12

Query Match 58.2%; Score 64; DB 10; Length 117;
 Best Local Similarity 52.9%; Pred. No. 3.7;
 Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

OY 8 KXXXXXXXXXSHPOFEK 24
 |:::|:::|:::|:::|:::|
 DB 101 KGNVSCPSAMSHPOFEK 117

RESULT 14
 US-09-977-137A-6

Sequence 6; Application US/09977137A
 Publication No. US20030104524A1
 GENERAL INFORMATION:

APPLICANT: Summers, Anne O.

APPLICANT: Cdquist, Jonathan
 TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
 Methods

FILE REFERENCE: 79-00

CURRENT APPLICATION NUMBER: US/09/977,137A

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,465

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 6

LENGTH: 118

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: chelon
 US-09-977-137A-6

Query Match 58.2%; Score 64; DB 10; Length 118;
 Best Local Similarity 52.9%; Pred. No. 3.7;
 Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

OY 8 KXXXXXXXXXSHPOFEK 24
 |:::|:::|:::|:::|:::|
 DB 102 KGNVSCPSAMSHPOFEK 118

RESULT 15

US-10-628-432-40

Sequence 40; Application US/10628432

Publication No. US20040142863A1

GENERAL INFORMATION:

APPLICANT: Wyeth

TITLE OF INVENTION: Modified ADAMTS4 molecules

FILE REFERENCE: AM101378

CURRENT APPLICATION NUMBER: US/10/628,432

CURRENT FILING DATE: 2003-07-29

NUMBER OF SEQ ID NOS: 53

SOFTWARE: Patentin version 3.1

SEQ ID NO 40

LENGTH: 845

TYPE: PRT

ORGANISM: Artificial

FEATURE:

OTHER INFORMATION: ADAMTS4 ASM with insertion
 US-10-628-432-40

Query Match 56.4%; Score 62; DB 16; Length 845;
 Best Local Similarity 44.4%; Pred. No. 67;
 Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY 7 EKKXXXXXXXXXSHPOFEK 24
 |:::|:::|:::|:::|:::|
 DB 510 DQLODFNIPQWSHPOFEK 527

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 27.3171 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ12
Perfect score: 110
Sequence: 1 wshpqlfexxxxxxxxxxwshpqlfex 24

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	58.2	21	4	US-09-809-517A-30
2	64	58.2	22	4	US-09-809-517A-33
3	64	58.2	117	4	US-09-977-137A-4
4	64	58.2	117	4	US-09-977-137A-5
5	64	58.2	117	4	US-09-977-137A-7
6	64	58.2	117	4	US-09-977-137A-8
7	64	58.2	117	4	US-09-977-137A-9
8	64	58.2	117	4	US-09-977-137A-10
9	64	58.2	117	4	US-09-977-137A-11
10	64	58.2	117	4	US-09-977-137A-12
11	64	58.2	118	4	US-09-977-137A-6
12	61	55.5	24	4	US-09-809-517A-31
13	61	55.5	25	4	US-09-809-517A-34
14	55	50.0	699	4	US-09-538-092-995
15	54	49.1	83	4	US-09-732-210-1454
16	54	49.1	512	3	US-09-413-814-8
17	54	49.1	521	3	US-08-504-048-9
18	53	48.2	10	4	US-09-809-517A-6
19	53	48.2	233	4	US-09-328-352-5498
20	52	47.3	2324	4	US-09-902-540-9732
21	51	46.4	8	3	US-08-948-097-2
22	51	46.4	8	3	US-09-382-950-7
23	51	46.4	8	3	US-09-382-7368-8
24	51	46.4	8	4	US-09-619-103-9
25	51	46.4	8	4	US-10-104-218-5
26	51	46.4	8	4	US-09-809-517A-9
27	51	46.4	86	4	US-09-198-452A-677

28	51	46.4	94	4	US-09-438-185A-640	Sequence 640, App
29	51	46.4	108	4	US-09-023-905A-32	Sequence 32, Appl
30	51	46.4	204	4	US-09-248-796A-16426	Sequence 16426, A
31	51	46.4	406	4	US-09-248-796A-16826	Sequence 16826, A
32	51	46.4	420	4	US-09-949-016-9364	Sequence 9364, Ap
33	51	46.4	476	4	US-09-949-016-10314	Sequence 10314, A
34	51	46.4	537	4	US-09-949-016-5954	Sequence 5954, Ap
35	50	45.5	158	4	US-09-621-976-4227	Sequence 4227, Ap
36	50	45.5	273	4	US-09-328-352-6417	Sequence 6417, Ap
37	50	45.5	356	4	US-09-198-452A-526	Sequence 526, App
38	50	45.5	487	4	US-09-602-777A-420	Sequence 420, App
39	50	45.5	527	4	US-09-600-985-1	Sequence 1, Appli
40	50	45.5	527	4	US-09-600-985-2	Sequence 2, Appli
41	50	45.5	527	4	US-09-600-985-3	Sequence 3, Appli
42	50	45.5	574	4	US-09-438-185A-489	Sequence 489, App
43	50	45.5	645	4	US-09-547-435-26	Sequence 26, Appl
44	50	45.5	771	4	US-09-547-435-28	Sequence 28, Appl
45	50	45.5	867	4	US-09-547-435-24	Sequence 24, Appl

ALIGNMENTS

```
RESULT 1
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on l
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
Query Match      58.2%; Score 64; DB 4; Length 21;
Best Local Similarity 52.9%; Pred. No. 0.026;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;
QY      8 KXXXXXXXXXSHPOFEK 24
Db      5 KDDDDKGAFWSHPOFEK 21
RESULT 2
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on l
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match          58.2%; Score 64; DB 4; Length 22;
Best Local Similarity 52.9%; Pred. No. 0.028;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      6 KDDDDKGAFWSHPOFEK 22

RESULT 3
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      101 KGNVSCPSAWSHPOFEK 117

RESULT 4
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
```

```

US-09-977-137A-5

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      101 KGNVSCPSAWSHPOFEK 117

RESULT 5
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      101 KGNVSCPSAWSHPOFEK 117

RESULT 6
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
```

```
Db      101 KGNVSCPSAMSHPOFEK 117
|:|||||
RESULT 7
US-09-977-137A-9
; Sequence 9, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 8
US-09-977-137A-10
; Sequence 10, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 9
US-09-977-137A-11
; Sequence 11, Application US/09977137A
```

```
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 10
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 11
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
```

```

; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match          58.2%; Score 64; DB 4; Length 118;
Best Local Similarity 52.9%; Pred. No. 0.29;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXWHPQFEK 24
Db 102 KGVSCPSAMSHPOFEK 118

RESULT 12
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 9914072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match          55.5%; Score 61; DB 4; Length 24;
Best Local Similarity 41.7%; Pred. No. 0.086;
Matches 10; Conservative 10; Mismatches 0; Indels 4; Gaps 1;

QY 5 OFEK-----XXXXXXXXXWHPQFEK 24
Db 1 EFQKLISEEDJNGAPWHPQFEK 24

RESULT 13
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 9914072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match          55.5%; Score 61; DB 4; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.091;
Matches 10; Conservative 10; Mismatches 0; Indels 4; Gaps 1;

QY 5 OFEK-----XXXXXXXXXWHPQFEK 24
Db 2 EFQKLISEEDJNGAPWHPQFEK 25

RESULT 14
US-09-538-092-995
; Sequence 995, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 995
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P23327
US-09-538-092-995

Query Match          50.0%; Score 55; DB 4; Length 699;
Best Local Similarity 36.8%; Pred. No. 67;
Matches 7; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 WHPQFEKXXXXXXXXXWHP 19
Db 83 WHPREKEDVSKETGH 101

RESULT 15
US-09-732-210-1454
; Sequence 1454, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitcanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 19.1057 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ13
Perfect score: 111
Sequence: 1 wshpqfexkxxxxxxxxwshpqfex 25

Scoring table:
BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR79:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	52.3	1113	2 T20004	hypothetical prote
2	57	51.4	73	2 S28978	srb protein - phag
3	56	50.5	759	2 D81657	general secretion
4	55	49.5	103	2 E90973	hypothetical prote
5	55	49.5	103	2 A85821	unknown protein en
6	55	49.5	103	2 C85713	unknown protein en
7	55	49.5	103	2 A90904	hypothetical prote
8	55	49.5	315	2 T23597	hypothetical prote
9	55	49.5	318	2 S76226	hypothetical prote
10	55	49.5	330	2 AC2071	hypothetical prote
11	55	49.5	379	2 T23592	hypothetical prote
12	54	48.6	260	2 AB2080	hypothetical prote
13	54	48.6	308	2 G86464	P12612.10 protein
14	54	48.6	414	2 G84311	hypothetical prote
15	54	48.6	625	2 G72420	oligopeptidase ABC
16	54	48.6	958	2 S62413	probable sulfatase
17	54	48.6	2150	1 S27802	zinc finger protei
18	54	48.6	2150	2 T19450	hypothetical prote
19	53	47.7	83	1 C42645	ribosomal protein
20	53	47.7	83	2 F81664	gelatin, ovarian
21	53	47.7	437	2 A31142	S-allele-associated
22	52.5	47.3	220	1 S22516	histidine rich cal
23	52.5	47.3	108	2 P95385	protein (imported
24	52	46.8	226	2 B83081	conserved hypothet
25	52	46.8	323	2 E88642	protein C54B4.5 (1
26	52	46.8	434	2 G95297	probable hydrolase
27	52	46.8	451	2 G01227	receptin - human
28	52	46.8	633	2 D72284	oligopeptide ABC
29	52	46.8	633	2 D72284	oligopeptide ABC

30	52	46.8	760	2 E71499	probable gen. secr
31	52	46.8	957	2 T10633	hypothetical prote
32	51.5	46.4	71	1 IHER1	high potential pro
33	51	45.9	109	2 E89873	cysteine proteina
34	51	45.9	225	2 A75425	hypothetical prote
35	51	45.9	318	2 B75382	conserved hypothet
36	51	45.9	656	2 T01573	earl protein - mal
37	51	45.9	664	2 A47012	phenol 2-monooxyge
38	51	45.9	754	2 B66592	general secretion
39	51	45.9	754	2 D72032	general secretion
40	51	45.9	1003	2 T28654	transposase - Pseu
41	51	45.9	3947	2 T52486	ferrichrome sidero
42	50.5	45.5	1078	2 S71162	DNA topoisomerase
43	50	45.0	86	2 C86570	S17 ribosomal prot
44	50	45.0	86	2 H72054	ribosomal protein
45	50	45.0	161	2 B42627	cement precursor p

ALIGNMENTS

```
RESULT 1
T20004
hypothetical protein C47E12.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Jul-2004
C:Accession: T20004
R:Colles, L.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19210
A:Accession: T20004
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1113 <WILL>
A:Cross-references: UNIPROT:Q27481; EMBL:Z68882; PDB:CAA93101.1; GSPDB:GN00022; CESP:C
A:Experimental source: clone C47E12
C:Genetics:
A:Gene: CESP:C47E12.5
A:Map position: 4
A:Introns: 65/3; 112/2; 213/3; 574/3; 968/3; 1070/2
C:Superfamily: ubiquitin-activating enzyme E1

Query Match
Query Length 52.3% Score 58; DB 2; Length 1113;
Best Local Similarity 34.8%; Pred. No. 37;
Matches 8; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

QY 2 SHPOFEKXXXXXXXXXWHPQFE 24
Db 355 SHVPEKSLVPEPFGIMDYAKFE 377

RESULT 2
S28978
srb protein - phage HK022
C:Species: phage HK022
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 03-Jul-2004
C:Accession: S28978; S16858
R:Atkinson, B.L.; Gottesman, M.E.
J. Mol. Biol. 227, 29-37, 1992
A:Title: The Escherichia coli ropB60 mutation blocks antitermination by coliphage HK022
A:Reference number: S28977; MUID:92595669; PMID:1522593
A:Accession: S28978
A:Molecule type: DNA
A:Residues: 1-73 <ATK>
A:Cross-references: UNIPROT:Q38575; EMBL:X60308; NID:G15753; PDB:CAA42853.1; PID:G57915
A:Genetics:
A:Gene: srb
A:Start codon: GTG

Query Match
Query Length 51.4%; Score 57; DB 2; Length 73;
Best Local Similarity 36.4%; Pred. No. 2.1;
Matches 8; Conservative 11; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      4 PQFEKXXXXXXXXXXWSHPQFEK 25
        | : : : : : : : : : : : :
Db      41 PDSKQQLLELTQYSHPPFEK 62
```

```

RESULT 3
D81657
General secretion pathway protein D TC0861 [imported] - Chlamydia muridarum (strain Nigg)
C|Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C|Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C|Accession: D81657
R|Read, T.D.; Brubham, R.C.; Shen, C.; Gall, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A|Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A|Reference number: AB1500; MUID:20150255; PMID:10684935
A|Accession: D81657
A|Status: Preliminary
A|Molecule type: DNA
A|Residues: 1-759 <TEXT>
A|Cross-references: UNIPROT:Q9PUG8; GB:AE002352; GB:AE002160; NID:g7190879; PIDD:AA839665
A|Experimental source: strain Nigg (MoPn)
C|Genetics:
C|Gene: TC0861

```

```

Query Match      50.5% Score 56; DB 2; Length 75;
Best Local Similarity 31.8% Pred. NO. 45;
Matches      7; Conservative 11; Mismatches 4; Indels 0; Gaps 0.

Qy      1 WSHPOFEKXXXXXXXXXXWSHPQ 22
      +-+-----+
      |||:|:|:|:|:|:|:|:|:|:|
      103 WSAEVSKEGKDPEEDYALNNHPE 124
      bb

```

hypochemical protein ECa2757 [imported] - *Escherichia coli* (strain O157:H7, substrain R
 C/Species: *Escherichia coli*
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C/Accession: E90973
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genot
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: E90973
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-103 <HAY>
 A/Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDD:BA86180.1; PTD:G13362225; GSPDB:G
 A/Experimental source: strain O157:H7, substrain RMD 0509952
 C/Genetics:
 C/Gene: ECa2757

Query Match	1	49.5%	Score 55;	DB 2;	Length 103;
Best Local Similarity		31.6%	Pred. NO. 6;		
Matches	6;	Conservative 12;	Mismatches 1;	Indels 0;	Gaps 0;
QY	7	EKXXXXXXXXXXSHPQPEK	25		
b	6	EMPERVNERNYGCMTHPDEYEK	24		

RESULT 5
 A85821
 unknown protein encoded within prophage CP-933v [imported] - *Escherichia coli* (strain O157:H7)
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C/Accession: A85821
 R/Perma.NT: Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 /Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: AB5821
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO-
A:Cross-references: UNIPROT:O8X4V0; GB:AE005174; NID:G12516136; PIDN:AMG57029.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3120

```

Query Match      49.5%  Score 55;  DB 2;  Length 103;
Best Local Similarity 31.6%  Pred. No. 6;
Matches 6;  Conservative 12;  Mismatches 1;  Indels 0;  Gaps 0;

QY 7 EKKXXXXXXXXXWSPQFEK 25
    | : : : : : : : : : : : : : : : :
Db 6 EMPVERNVGGCWTBPEYEK 24

```

RESULT 6
 C85713
 unknown protein encoded within prophage CP_9330 [imported] - Escherichia coli (strain O1
 C)Species: Escherichia coli
 C|Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C|Accession: C85713
 R|Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A|Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A|Reference number: A85480; MUID:21074935; PMID:11206551
 A|Accession: C85713
 A|Status: preliminary
 A|Molecule type: DNA
 A|Residues: 1*103 <STO>
 A|Cross-references: UNIPROT:O8X549; GB:AE005174; NID:G12515046; PIDN:AA056167.1; GSPDB:G
 A|Experimental source: strain O157:H7, substrain EDL933
 A|Genetics:
 A|Gene: Z2097

```

Query Match      49.5%; Score 55; DB 2; Length 103;
Best Local Similarity 31.6%; Pred. No. 6;
Matches 6; Conservative 12; Mismatches 1; Indels 0; Gaps 0;

Qy 7 EKXXXXXXXXXXSMHPQEK 25
    | : : : : : : : : : : : :
Db 6 EMPVVRDGYGYMTHPREYK 24

```

```

RESULT 7
A:90904
hypothetical protein ECs2201 [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 09-Jul-2004
C:Accession: A90904
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno-
A:Reference number: A96629; MUID:21156231; PMID:11258796
A:Accession: A90904
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAV>
A:Cross-References: UNIPROT:O8X49, GB:BA000007, PIDD:BAB35624.1; PID:g13361667, GSPDB:C
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs2201

```

Query Match 49.5%; Score 55; DB 2; Length 103;
Best Local Similarity 31.6%; Pred. No. 6;
Matches 6; Conservative 12; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 92.8862 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ13
Perfect score: 111
Sequence: 1 webpqlfexxxxxxxxxwebpqlfex 25

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	58.6	743	2	Q8CH90
2	63	56.8	270	2	Q8BLU4
3	63	56.8	663	2	Q8U219
4	62	55.9	455	2	Q6MKM3
5	61	55.0	421	2	Q9U6S0
6	60	54.1	704	2	Q8CJD2
7	58	52.3	212	2	Q7TSS9
8	58	52.3	1113	2	Q27481
9	57	51.4	73	2	Q38575
10	57	51.4	406	2	Q7MMZ9
11	57	51.4	406	2	Q8DFE9
12	57	51.4	410	2	Q64KA0
13	56	50.5	207	2	Q90B81
14	56	50.5	759	2	Q9P0G8
15	55.5	50.0	323	2	Q8LRT5
16	55	49.5	103	2	Q8X549
17	55	49.5	103	2	Q8X4V0
18	55	49.5	107	2	Q6Z315
19	55	49.5	315	2	Q9XU55
20	55	49.5	318	2	P74388
21	55	49.5	330	2	Q8YV60
22	55	49.5	379	2	Q45680
23	55	49.5	410	2	Q64K95
24	55	49.5	410	2	Q64K98
25	55	49.5	410	2	Q64K99
26	55	49.5	411	2	Q64K96
27	55	49.5	411	2	Q64K97
28	55	49.5	499	2	Q7Q4F7
29	55	49.5	508	1	MAR MARSC
30	55	49.5	508	2	Q9BBG0
31	55	49.5	508	2	Q9BBG3

32	55	49.5	509	2	Q9VMT4	Q9vmt4 drosophila
33	54	48.6	144	2	Q6F91	Q6f91 human immun
34	54	48.6	251	2	Q9FZU5	Q9fzu5 neisseria m
35	54	48.6	260	2	Q8YU9	Q8yu9 anabaena sp
36	54	48.6	308	2	Q9FX16	Q9fx16 arabidopsis
37	54	48.6	414	2	Q9HPR8	Q9hpr8 halobacteri
38	54	48.6	625	2	Q9WXS6	Q9wxs6 thermotoga
39	54	48.6	638	2	Q8D796	Q8d796 vibrio vuln
40	54	48.6	668	2	Q7ME40	Q7me40 vibrio vuln
41	54	48.6	698	2	Q9P8D7	Q9p8d7 candida clo
42	54	48.6	958	1	YA7B_SCHPO	Q99764 schizosach
43	54	48.6	1372	2	Q8Z2M8	Q8z2m8 enterococcu
44	54	48.6	2150	1	SDC3 CAEBL	P34706 caenorhabdi
45	53.5	48.2	843	2	Q6CKE4	Q6cke4 kluyveromyc

ALIGNMENTS

RESULT 1									
ID	Q8CH90	PRELIMINARY;	PRT;	743	AA.				
AC	Q8CH90;								
DT	01-MAR-2003 (TREMBLrel. 23, Created)								
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)								
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)								
DE	Soluble guanylyl cyclase alpha 2 subunit E219G mutant.								
OS	Rattus norvegicus (Rat).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.								
OX	NCBI_Taxid=10116;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	TISSUE=Kidney;								
RA	Nakamura I., Yao Y., Suzuki N.;								
RL	Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.								
DR	EMBL; AB097860; BAC44887.1; -								
DR	HSSP; P10803; IASZ.								
DR	GO; GO:0004383; F:guanylate cyclase activity; IEA.								
DR	GO; GO:0016829; F:lyase activity; IEA.								
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.								
DR	InterPro; IPR01054; G:G cyclase.								
DR	InterPro; IPR009080; tRNAasn_1a_bind.								
DR	Pfam; PF00211; Guanylate_cyc; 1.								
DR	SMART; SM00044; CYC; 1.								
DR	PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.								
DR	PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.								
KW	Lyase.								
SQ	SEQUENCE 743 AA; 83251 MW; ACP5C53E0982813A CRC64;								
Query Match									
Best Local Similarity 58.6%; Score 65; DB 2; Length 743;									
Matches 9; Conservative 10; Mismatches 1; Indels 0; Gaps 0;									
QY	6	PEKXXXXXXXXXSHPOFEK	25						
DB	724	FLKRTSLVPGSMHPOFEK	743						
RESULT 2									
ID	Q9BLU4	PRELIMINARY;	PRT;	270	AA.				
AC	Q9BLU4;								
DT	01-JUN-2001 (TREMBLrel. 17, Created)								
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)								
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)								
DE	Hypothetical protein L490.02.								
GN	Name=L490.02;								
OS	Leishmania major.								
OC	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.								
OC	NCB1_Taxid=5664;								
OX	NCBI_Taxid=5664;								
RN	[1]								
RP	SEQUENCE FROM N.A.								

RC STRAIN=Friedlin;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzaden A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the *Leishmania* major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friedlin;
 RA Clariom L., Tosato V., Bruschi C.V., Ivens A.C., Quail M.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL512295; CAC24652.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 270 AA; 29661 MW; 60829CD354CFB36 CRC64;

Query Match 56.8%; Score 63; DB 2; Length 270;
 Best Local Similarity 34.8%; Pred. No. 9.8;
 Matches 8; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

Oy 3 HPOFEKXXXXXXXXXXSHPOFK 25
 Db 18 HPOFOAVNVGQRVVAHPTFK 40

RESULT 3

ID 08J219 PRELIMINARY; PRT; 663 AA.
 AC 08J219;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Aminopectidase C.
 GN Name=apcc;
 OS *Aspergillus niger*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N400;
 RX MEDLINE=22459136; PubMed=12571053;
 RA DOI=10.1128/AEM.69.2.1246-1250.2003;
 RA Baasen D.E.J.W., Dekker P.J.T., Schaap P.J.;
 RT "Aminopectidase C of *Aspergillus niger* is a novel phenylalanine
 RT aminopectidase";
 RL Appl. Environ. Microbiol. 69:1246-1250(2003).
 DR EMBL; AJ316576; CAC87723.1; -
 DR MEROPS; S09.057; -
 DR GO; GO:0004177; F:aminopectidase activity; IEA.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR008262; Lipase AS.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser_esters.
 DR InterPro; IPR01042; ToB_C.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
 DR PROSITE; PS00300; SRP54; UNKNOWN_1.
 KW Aminopectidase.
 SQ SEQUENCE 663 AA; 72517 MW; C4EA59DA9743B545 CRC64;

Query Match 56.8%; Score 63; DB 2; Length 663;
 Best Local Similarity 31.8%; Pred. No. 27;
 Matches 7; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

Oy 1 WSHPOFEKXXXXXXXXXXSHPO 22
 Db 176 YSHPTSPAGDRVSWIOWNHPE 197

RESULT 4

OGMKM3

ID 06MKM3 PRELIMINARY; PRT; 455 AA.
 AC 06MKM3;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DE Hypothetical protein precursor.
 GN OrderedLocustNames=Bd2365;
 OS *Bdellovibrio bacteriovorus*.
 OC Bacteria; Proteobacteria; Delaproteobacteria; *Bdellovibrionales*;
 OC *Bdellovibrionaceae*; *Bdellovibrio*.
 OX NCBI_TaxID=959;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
 RX PubMed=14752164; DOI=10.1126/science.1093027;
 RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,
 RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
 RA Sockett R.E., Schuster S.C.;
 RT "A predator unmasked: life cycle of *Bdellovibrio bacteriovorus* from a
 RT genomic perspective";
 RL Science 303:689-692(2004).
 DR EMBL; BX842652; CAE80184.1; -
 KW Complete proteome.
 FT SIGNAL
 SQ SEQUENCE 455 AA; 51733 MW; 94F51E24304BDF9 CRC64;

Query Match 55.9%; Score 62; DB 2; Length 455;
 Best Local Similarity 30.4%; Pred. No. 24;
 Matches 7; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

Oy 1 WSHPOFEKXXXXXXXXXXSHPOF 23
 Db 311 WTHPTFEDAVVSPFVETNSYF 333

RESULT 5

ID 09U6S0 PRELIMINARY; PRT; 421 AA.
 AC 09U6S0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Calreticulin precursor.
 GN Name=calret;
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Susan J.M., Just M.L., Lennarz W.J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1-SIMILARITY: Belongs to the calreticulin family.
 DR EMBL; AF177915; AAD55725.1; -
 DR HSSP; P18418; IK9C.
 DR GO; GO:0005783; C:endoplasmic reticulum; IEA.
 DR GO; GO:0005514; F:calcium ion storage activity; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR001580; Calret/calnex.
 DR InterPro; IPR009169; Calreticulin.
 DR InterPro; IPR009033; Calret calnex P.
 DR InterPro; IPR008985; Cona_like_1ec_g1.
 DR InterPro; IPR000886; ER_target_S.
 DR Pfam; PF00262; Calreticulin; 1.
 DR PRINTS; PR00626; CALRETICULIN.1.
 DR PROSITE; PS00804; CALRETICULIN_2; 1.
 DR PROSITE; PS00805; CALRETICULIN_REPEAT; 3.
 DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
 KW Chapterone; Endoplasmic reticulum; Signal.
 FT SIGNAL
 SQ SEQUENCE 421 AA; 46173 MW; 94F51E24304BDF9 CRC64;

FT	CHAIN	20	421	calreticulin.
50	SEQUENCE	421 AA;	48822 MW;	173C664F59F41P93 CRC64;
Query Match				
	Best Local Similarity	30.4%;	Score 61;	DB 2; Length 421;
	Matches	7; Conservative	12; Mismatches	4; Indels
				0; Gaps
Qy	1	MSHPQFKXXXXXXXXXMSHPQF 23		
Db	288	VWHPETENSEYEADENLYSPSF 310		
RESULT 6				
Q8CUD2	PRELIMINARY;	PRT;	704 AA.	
ID	Q8CUD2			
AC	Q8CUD2;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Guanylyl cyclase alpha 1 subunit.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Nakamura I., Suzuki N.,			
RL	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB096020; BAC24016.1; -.			
DR	HSSP; P30803; IAZS.			
DR	GO; GO:0004383; F:guanylate cyclase activity; IEA.			
DR	GO; GO:0016828; F:lyase activity; IEA.			
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.			
DR	InterPro; IPR01054; G_cyclase.			
DR	Pfam; PF00211; Guanylate_cyc; 1.			
DR	SMART; SM00044; CYC; 1.			
DR	PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.			
DR	PROSITE; PS01025; GUANYLATE_CYCLASES_2; 1.			
KW	Lyase.			
SQ	SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;			
Query Match				
	Best Local Similarity	47.1%;	Score 60;	DB 2; Length 704;
	Matches	8; Conservative	9; Mismatches	0; Indels
				0; Gaps
Qy	9	XXXXXXXXXXMSHPQFK 25		
Db	688	GVDLVPRGSMWHPQFK 704		
RESULT 7				
Q7TS89	PRELIMINARY;	PRT;	212 AA.	
ID	Q7TS89			
AC	Q7TS89;			
DT	01-OCT-2003 (TREMBLrel. 25, Created)			
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Hypothetical protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Mammary gland;			
RX	MELN1=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klauber R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buecaw K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Datchenko L., Marusich K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stephenson M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,			
RA	Brownstein M.C., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,			

Ra Ra S.S., Longwellano N.A., Peeters G.J., Abramson R.D., Mullaly S.J.,
Ra Beak S.A., McWnan P.J., McEnern K.J., Malek J.A., Gunaratne P.H.,
Ra Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Ra Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Ra Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
Ra Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Ra Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Ra Krzywnski M.T., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
Ra Jones S.J., Marra M.A.;
Rt "Generation and initial analysis of more than 15,000 full-length human
Rt and mouse cDNA sequences.";
Rt Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Strauberg R.;
RL Submitted (MAY-2003) to the EMBL/Genbank/DBD databases.
DR EMBL: BC052931; ANH52931.1; -
DR GO: GO:0005634; C:nucleus; IEA.
DR InterPro: IPR004865; Spi100.
DR Pfam: PF031172; Spi100; 1.
KW Hypothetical protein.
SQ SEQUENCE 212 AA; 24759 MW; 817B5D56E1A6FE36 CRC64;

Query Match 52.3%; Score 58; DB 2; Length 212;
Best Local Similarity 30.4%; Pred. No. 36;
Matches 7; Conservative 12; Mismatches 4; Indels 0; Gaps

Oy 3 HPQEKXXXXXXXXXXSHPOPEK 25
Db |||.:|.....:||||||:
130 HPSYEGTNDLCTHEKAQPCFLK 152

RESULT 8	027481	PRELIMINARY;	PRT; 1113 AA.
ID	Q27481		
AC	Q27481;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Hypotheetical protein C47E12.5.		
OS	Name=uba-1; ORFNames=C47E12.5;		
ON	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Br1etrol N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	none;		
RT	"genome sequence of the nematode C.elegans: A platform for		
RT	investigating biology."		
RL	Science 282:2012-2018 (1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Br1etrol N2;		
RA	Collee L.;		
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; Z68882; CAA93101.1; -.		
DR	PIR; T20004; T20004.		
DR	HSSP; Q8TBC4; INGV.		
DR	Wormbase; WBGene00006699; uba-1.		
DR	Wormped; C47E12.5; CE05449.		
DR	GO; GO:0003824; F: catalytic activity; IEA.		
DR	GO; GO:0004839; F: ubiquitin activating enzyme activity; IEA.		
DR	GO; GO:0006112; F: ubiquitin cycle; IEA.		
DR	InterPro; IPR009036; Moeb.		
DR	InterPro; IPR000205; NAD BS.		
DR	InterPro; IPR000594; Th1F domain.		
DR	InterPro; IPR000127; UBacC_repeat.		
DR	InterPro; IPR000011; Ubqin-activ enz.		

```

DR Pfam: PF00899; Trif; 2.
DR Pfam: PF02334; UBACT; 2.
DR PRINTS; PR01849; UBIQUITINACT.
DR TIGRFAMs; TIGR01408; Ubel; 1.
DR PROSITE; PS00865; UBIQUITIN_ACTIVAT_2; 1.
DR KEGG; K01113; UBIQUITIN.
SQ SEQUENCE 1113 AA; 124083 MW; 8D2198A97E4066CC CRC64;

Query Match
Best Local Similarity 52.3%; Score 58; DB 2; Length 1113;
Matches 8; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SHPOFEKXXXXXXXXXWSHPQ 24
Db 355 SHVPEKSLVPEFGIMDYAKE 377

RESULT 9
ID Q38575; PRELIMINARY; PRT; 73 AA.
AC Q38575;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Srp.
OS Bacteriophage HK022.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=10742;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92395669; PubMed=1522593;
RA Atkinson B.L., Gottesman M.E.,
RT "The E. coli IP080 mutation blocks antitermination by coliphage HK022
RT Q-function."
RT J. Mol. Biol. 227:29-37(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20328598; PubMed=10860721; DOI=10.1006/jmbi.2000.3729;
RA Junia R.U., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,
RT Hendrix R.W.,
RT "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic
RT mosaicism in the lambdaoid bacteriophages."
RT J. Mol. Biol. 299:27-51(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Junia R., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,
RA Hendrix R.W.,
RT Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; X60308; AAA42853.1; -.
DR EMBL; AF069308; AAF30374.1; -.
DR PIR; S28978; S28978.
SQ SEQUENCE 73 AA; 8302 MW; AAC13F263A3F42CF CRC64;

Query Match
Best Local Similarity 51.4%; Score 57; DB 2; Length 73;
Matches 8; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

Qy 4 POFEXKXXXXXXXXXWSHPQ 25
Db 41 PDSKXQILIELTQXSHPEFEK 62

RESULT 10
ID Q7MMZ9; PRELIMINARY; PRT; 406 AA.
AC Q7MMZ9;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE 2-poly(phenyl-6-methoxyphenol) hydroxylase.
OS OrderedLocustNames=VVO918;
OS Vibrio vulnificus (strain VVO16).

```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14656965; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Teal H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Teal S.-F.,
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen."
RL Genome Res. 13:2577-2587(2003).
DR EMBL; AP005333; BAC93682.1; -.
DR GO; GO:0004497; P:monooxygenase activity; IEA.
DR GO; GO:0006725; P:aromatic compound metabolism; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000733; Flav_monooxygenase.
DR InterPro; IPR00205; NAD BS.
DR InterPro; IPR003042; Rng_monooxygenase.
DR InterPro; IPR010971; Ubl_Ohaases.
DR Pfam; PF01360; Monooxygenase; 1.
DR PRINTS; PR00420; RINGMONOXGNASE.
DR TIGRFAMs; TIGR01988; Ubl_Ohaases; 1.
DR Complete proteome.
SQ SEQUENCE 406 AA; 45816 MW; 20A727877811CC98 CRC64;

Query Match
Best Local Similarity 51.4%; Score 57; DB 2; Length 406;
Matches 6; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHPQ 22
Db 91 WQHQTATRCVCPYKRLTEWHEP 112

RESULT 11
ID Q8DPE9; PRELIMINARY; PRT; 406 AA.
AC Q8DPE9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE 2-poly(phenyl-6-methoxyphenol) hydroxylase.
OS OrderedLocustNames=VV10265;
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.,
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016797; AA008799.1; -.
DR GO; GO:0004497; P:monooxygenase activity; IEA.
DR GO; GO:0006725; P:aromatic compound metabolism; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000733; Flav_monooxygenase.
DR InterPro; IPR003042; Rng_monooxygenase.
DR InterPro; IPR010971; Ubl_Ohaases.
DR Pfam; PF01360; Monooxygenase; 1.
DR PRINTS; PR00420; RINGMONOXGNASE.
DR TIGRFAMs; TIGR01988; Ubl_Ohaases; 1.
DR Complete proteome.
SQ SEQUENCE 406 AA; 45880 MW; 3741FDC327ED37F CRC64;

Query Match
Best Local Similarity 51.4%; Score 57; DB 2; Length 406;
Matches 6; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHPQ 22

```

Db 91 WQHIOATRCVCPKRLTWHPE 112

RESULT 12

064KAO PRELIMINARY; PRT; 410 AA.
AC 064KAO;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Calreticulin.
OS Amblyomma brasiliense.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Amblyomma.
OK NCBI_TaxID=257691;
RN (1)
RP SEQUENCE FROM N.A.
RA Xu G., Fang Q.Q., Kelrane J.E., Durden L.A.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY395247; AAR2933.1; -.
SQ SEQUENCE 410 AA; 47241 MW; B6235DEFA8D378B1 CRC64;

Query Match 51.4%; Score 57; DB 2; Length 410;

Best Local Similarity 22.7%; Pred. No. 1e+02;
Matches 5; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHPQ 22

Db 288 WVHPEIDNPEYADPRLVHPE 309

RESULT 13

Q90B81 PRELIMINARY; PRT; 207 AA.
AC Q90B81;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Nef.
OS Name=nef;
GN Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OK NCBI_TaxID=11676;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=21637600; PubMed=11779357; DOI=10.1089/08992220175342103;
RA Maeshlsh T., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,
RA Abdool-Karim S., Williamson C., Gray C.M.;
RT "Conserved domains of subtype C nef from South African HIV type 1-
RT infected individuals include cytotoxic T lymphocyte epitope-rich
RT regions."
RL Aids Res. Hum. Retroviruses 17:1681-1687(2001).
RN (12)
RP SEQUENCE FROM N.A.
RA Maeshlsh T.N., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,
RA Abdool-Karim S., Williamson C., Gray C.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF397538; AAK98477.1; -.
DR HSP; P04324; 1ZEC.
DR GO; GO:0005525; F:GTP binding; IEA.
DR Pfam; PF00469; F-protein 1.
DR Pfdom; PD000031; HIV_Nef; 1.
KW Aids; Lipoprotein; Myristate.
SQ SEQUENCE 207 AA; 23626 MW; FEC73E6A5D79D78 CRC64;

Query Match 50.5%; Score 56; DB 2; Length 207;

Best Local Similarity 28.6%; Pred. No. 66;
Matches 6; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

Qy 5 QFEKXXXXXXXXXWSHPQEK 25

Db 185 EFDSSLARHNLGRMKRFEYK 205

RESULT 14

Q9PJG8 PRELIMINARY; PRT; 759 AA.
AC Q9PJG8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE General secretion pathway protein D.
GN Name=gspD; OrderedLocustNames=TC0861;
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OK NCBI_TaxID=83560;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / Nig9;
RX MEDLINE=20150255; PubMed=10684935; DOI=10.1093/nar/28.6.1397;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey B.K., Peterson J.D., Uppertack T.R., Berry K.J.,
RA Bass S., Linher K.D., Weidman J.F., Khouri H.W., Craven B., Bowman C.,
RA Dodson R.J., Gwinn M.L., Nelson W.C., Deboy R.T., Kolonay J.F.,
RA McClarty G., Salzberg S.L., Eisen J.A., Fraser C.M.;
RL "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae A839."
RT Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- SIMILARITY: Belongs to the GSP D family.
DR EMBL; AB002352; AAF39657.1; -.
DR PIR; D81657; D81657.
DR TIGR; TC0861; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0009306; P:protein secretion; IEA.
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR004846; GSPD/IIPprotein.
DR InterPro; IPR001220; Lectin legB.
DR InterPro; IPR005644; NoIw-like.
DR Pfam; PF00263; Secretin_1.
DR Pfam; PF03958; Secretin_N; 2.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PROSITE; PS00307; LECTIN LEGUME_BETA; UNKNOWN_1.
KW Complete proteome; Transport.
SQ SEQUENCE 759 AA; 83771 MW; 7D5DDF55ED3FBA31 CRC64;

Query Match 50.5%; Score 56; DB 2; Length 759;

Best Local Similarity 31.8%; Pred. No. 2.8e+02;
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHPQ 22

Db 103 WSAEVSERKGDPEYALWNHPE 124

RESULT 15

Q8LTT5 PRELIMINARY; PRT; 323 AA.
AC Q8LTT5;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE 5' to 3' exonuclease-like protein.
OS Vibrio phage VPV262.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OK NCBI_TaxID=194802;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=2267917; PubMed=12781722;
RA Hardies S.C., Comeau A.M., Server P., Suttle C.A.;
RT "The complete sequence of marine bacteriophage VPV262 infecting vibrio
RT parahaemolyticus indicates that an ancestral component of a T7 viral
RT supergroup is widespread in the marine environment."
RL Virology 310:359-371(2003).

DR EMBL: AY095314; AAM28378.1; -
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0004527; F:exonuclease activity; IEA.
DR InterPro: IPR008918; 5_3_exo_C.
DR InterPro: IPR000513; Exo_N_I.
KM Exonuclease.
SQ SEQUENCE 323 AA; 37725 MW; 81CB297839B33951 CRC64;

Query Match 50.0%; Score 55.5; DB 2; Length 323;
Best Local Similarity 34.8%; Pred. No. 1.3e+02;
Matches 8; Conservative 12; Mismatches 2; Indels 1; Gaps 1;

QY 3 HPOFEKXXXXXXXXXWGHPOFEK 25
|||:||||:||||:|:|:
Db 78 HPEFEKRVADVDIHFWIE-EFDK 99

Search completed: March 2, 2005, 12:44:28
Job time : 94.8862 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 112.602 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ13
Perfect score: 111
Sequence: 1 wbnpgfexkxxxxxxwbnpgfexk 25

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq15Dec04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20048:*
8: geneseqp20058:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	89.6	24	ABP60362	Abp60362 Streptococci
2	96	86.5	35	ABP60369	Abp60369 Streptococci
3	95.5	86.0	36	ABP60370	Abp60370 Streptococci
4	71	64.0	845	ADS20243	Human agg
5	68	61.3	205	AAW93967	AAW93967 Plasmid p
6	68	61.3	254	AAW93969	AAW93969 Plasmid p
7	68	61.3	254	AAW93969	AAW93969 Plasmid p
8	68	61.3	396	AAW93966	AAW93966 Plasmid p
9	68	61.3	396	AAW93966	AAW93966 Plasmid p
10	68	61.3	396	AAW93966	AAW93966 Plasmid p
11	68	61.3	448	AAW93966	AAW93966 Plasmid p
12	68	61.3	448	AAW93966	AAW93966 Plasmid p
13	68	61.3	448	AAW93966	AAW93966 Plasmid p
14	68	61.3	448	AAW93966	AAW93966 Plasmid p
15	68	61.3	448	AAW93966	AAW93966 Plasmid p
16	68	61.3	448	AAW93966	AAW93966 Plasmid p
17	68	61.3	448	AAW93966	AAW93966 Plasmid p
18	68	61.3	448	AAW93966	AAW93966 Plasmid p
19	68	61.3	448	AAW93966	AAW93966 Plasmid p
20	68	61.3	448	AAW93966	AAW93966 Plasmid p
21	68	61.3	448	AAW93966	AAW93966 Plasmid p
22	68	61.3	448	AAW93966	AAW93966 Plasmid p
23	68	61.3	448	AAW93966	AAW93966 Plasmid p
24	68	61.3	448	AAW93966	AAW93966 Plasmid p
25	68	61.3	448	AAW93966	AAW93966 Plasmid p

26	63	56.8	269	8	ADR28054	Adr28054 NPB polyp
27	63	56.8	269	8	ADR28054	Adr28054 NPB polyp
28	63	56.8	273	8	ADO25157	Ado25157 Melanoma
29	63	56.8	274	8	ADO25157	Ado25157 Melanoma
30	63	56.8	275	8	ADO25154	Ado25154 Melanoma
31	63	56.8	275	8	ADO25155	Ado25155 Melanoma
32	63	56.8	276	8	ADO25149	Ado25149 Melanoma
33	63	56.8	277	8	ADO25150	Ado25150 Melanoma
34	63	56.8	277	8	ADO25152	Ado25152 Melanoma
35	63	56.8	278	8	ADO25151	Ado25151 Melanoma
36	63	56.8	281	8	ADO25156	Ado25156 Melanoma
37	63	56.8	288	8	ADR28055	Adr28055 NPB polyp
38	63	56.8	485	8	ADS20249	AdS20249 Furin-pro
39	63	56.8	663	5	AAU79048	Aau79048 Aspergillus
40	63	56.8	697	8	ADS20227	AdS20227 Human agg
41	62	55.9	117	5	AAU97558	Aau97558 Synthetic
42	62	55.9	117	5	AAU97553	Aau97553 Synthetic
43	62	55.9	117	5	AAU97557	Aau97557 Synthetic
44	62	55.9	117	5	AAU97559	Aau97559 Synthetic
45	62	55.9	117	5	AAU97555	Aau97555 Synthetic

ALIGNMENTS

RESULT 1	ABP60362	standard; peptide; 24 AA.
ID	ABP60362	standard; peptide; 24 AA.
AC	ABP60362;	
XX		
DT	28-MAR-2003	(first entry)
DE	Streptavidin binding peptide SEQ ID NO 3.	
XX		
KW	Streptavidin; protein chip; microtitre plate; detection.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 9.16	
FT		/label= unknown
XX		
PN	DE10113776-Al.	
XX		
PD	02-OCT-2002.	
XX		
PF	21-MAR-2001; 2001DE-01013776.	
XX		
PR	21-MAR-2001; 2001DE-01013776.	
XX		
PA	(BIOA-) INST BIOANALYTIC GMBH GOETTINGEN.	
XX		
PI	Schmidt T;	
XX		
DR	WPI; 2003-031166/03.	
XX		
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
PT		
PS	Disclosure; Page 4; 18pp; German:	
XX		
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily	

CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX
 SO Sequence 24 AA:

Query Match 89.6%; Score 99.5; DB 6; Length 24;
 Best Local Similarity 96.0%; Pred. No. 6e-06;
 Matches 24; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 25
 |||||
 DB 1 WSHPOFEK-XXXXXXXXXXWSHPQFEK 24

RESULT 2
 ABP60369

ID ABP60369 standard; peptide; 35 AA.

AC ABP60369;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 10.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 9..27 /label= unknown

FT /note= "optionally deleted for 1-15 residues"

DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

DR WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant
 PT protein, comprises at least two high-affinity streptavidin-binding
 PT modules.

PS Claim 7; Page 16; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX

Sequence 35 AA;

Query Match 86.5%; Score 96; DB 6; Length 35;

Best Local Similarity 71.4%; Pred. No. 3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 25
 |||||
 DB 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 35

RESULT 3

ID ABP60370 standard; peptide; 36 AA.

AC ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

FT Key Location/Qualifiers

FT Region 9..28 /note= "GGGS repeats 2-5 optionally absent, residues 13-28"

FT Region 9..12 /label= GGGS_repeat

DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

DR WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant
 PT protein, comprises at least two high-affinity streptavidin-binding
 PT modules.

PS Claim 8; Page 16; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX

Sequence 36 AA;

Query Match 86.0%; Score 95.5; DB 6; Length 36;
 Best Local Similarity 44.4%; Pred. No. 3.7e-05;
 Matches 16; Conservative 9; Mismatches 0; Indels 11; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 25
 |||||
 DB 1 WSHPOFEKGGSGGGSGGGSGGGSGWSHPQFEK 36

RESULT 4
 ADS20243
 ID ADS20243 standard; protein; 845 AA.
 XX
 AC ADS20243;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human aggrecanase ADAMTS4 mutant E362Q protein with inserted Strep tag.
 XX
 KW ADAMTS4; a disintegrin-like and metalloprotease;
 KW thrombospondin type 1 motif 4; reprolysin; zinc metalloprotease;
 KW aggrecanase; osteoarthritic; antiinflammatory; antirheumatic;
 KW cytostatic; osteoarthritis; glioma; cancer; inflammatory joint;
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
 KW human; enzyme; chromosome 1q21-q23; Strep tag; mutant; mutain.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Unidentified.
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 362
 FT Misc-difference /note="Wild-type Gln substituted by Gln"
 FT Misc-difference 520..527
 FT /note="Strep tag peptide 2 inserted"
 FT
 XX
 PN WO2004011637-A2.
 XX
 PD 05-FEB-2004;
 XX
 PF 29-JUL-2003; 2003WO-US023484.
 XX
 PR 29-JUL-2002; 2002US-0398721P.
 XX
 PA (AMHP) WYETH.
 PA (CORC) CORCORAN C J.
 PA (FLAN) FLANNERY C R.
 PA (ZENG) ZENG W.
 PA (RACI) RACIE L A.
 PA (MCDO) MCDONAGH T.
 PA (FREE) FREEMAN B A.
 PA (GEOR) GEORGIADIS K E.
 PA (LAVA) LAVALLIE E R.
 XX
 PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
 PI Freeman BA, Georgiadis KE, Lavallie ER;
 XX
 DR WPI; 2004-143860/14.
 XX
 PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
 PT stability useful for identifying inhibitors of the enzyme activity for
 PT treating aggrecanase-associated conditions, including osteoarthritis.
 XX
 ES Claim 9; SEQ ID NO 40; 117pp; English.
 XX
 CC The invention relates to a novel isolated, modified ADAMTS4 (a
 CC disintegrin-like and metalloprotease (reprolysin type) with
 CC thrombospondin type 1 motif 4) protein with improved stability compared
 CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
 CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
 CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc
 CC metalloproteases and include aggrecanases amongst their members. The
 CC protein of the invention demonstrates osteoarthritic, antiinflammatory,
 CC antirheumatic, antirheumatic and cytostatic activities and may be useful
 CC for treating aggrecanase-associated conditions, including osteoarthritis,
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
 CC arthritis, periodontal disease and Crohn's disease. The current sequence
 CC is that of the human aggrecanase ADAMTS4 mutant E362Q protein with
 CC inserted Strep tag of the invention.
 XX

SQL Sequence 845 AA;
 Query Match 64.0%; Score 71; DB 8; Length 845;
 Best Local Similarity 50.0%; Pred. No. 5.8;
 Matches 9; Conservative 9; Mismatches 0; Indels 0; Gaps 0;
 OY 1 WSHPOFEKXXXXXXXXXW 18
 |||||:||||:||||:||||:
 DB 520 WSHPOFEKAGGMPGMPW 537
 XX
 RESULT 5
 AAW93967
 ID AAW93967 standard; protein; 205 AA.
 XX
 AC AAW93967;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Plasmid pBBP21 protein fragment.
 XX
 KW Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;
 KW tissue-specific surface molecule; tumour-specific surface molecule;
 KW tumour imaging; bilin binding protein; DbC protein.
 XX
 OS Synthetic.
 OS
 FT Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= signal_peptide
 FT Protein 22..205
 FT /label= fusion_peptide
 FT /note="This fusion peptide is constructed from bilin
 FT binding protein and a strep-tag II region."
 FT
 XX
 PN WO9916873-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 25-SEP-1998; 98WO-DE002898.
 XX
 PR 26-SEP-1997; 97DE-01042706.
 XX
 PA (SKER) SKERRA A.
 XX
 PI Skerra A, Beese G, Schmidt F, Stibora T;
 XX
 DR WPI; 1999-255093/21.
 XX
 DR N-PSDB; AAX24100.
 XX
 PT New anticalins produced by mutation of lipocalin family protein.
 XX
 PS Disclosure; Page 70-72; 80pp; German.
 XX
 CC This invention describes the preparation of anticalins from polypeptides
 CC of the lipocalin family by mutation of amino acids in the region of the 4
 CC peptide loops (at one end of the cylindrical "leaflet" structure).
 CC Anticalins can bind to a predetermined ligand with determinable affinity.
 CC Anticalins or fusions of anticalins, are used: (a) when immobilized, for
 CC fusions or conjugates). Typically, they bind to tissue- or tumour-specific
 CC surface molecules and can be used for tumour imaging or directly for
 CC tumour therapy. Mutations can be introduced into lipocalin polypeptides
 CC more easily than into antibodies, since lipocalin polypeptides have only
 CC 4 (contrast 6) sequence segments. The peptide loops can tolerate amino
 CC acid changes without significant effect on folding. Anticalins have high
 CC specific affinity, comparable with that for antibodies. This sequence
 CC represents a fusion polypeptide used in the method of the invention
 XX
 SQ Sequence 205 AA;
 Query Match 61.3%; Score 68; DB 2; Length 205;
 Best Local Similarity 39.1%; Pred. No. 2.2;

[illegible]

```

SQ      Sequence 254 AA;                               61.3%; Score 68; DB 2; Length 254;
Query Match Similarity 39.1%; Pred. No. 2.9;
Best Local Similarity 39.1%; Pred. No. 2.9;
Matches          9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

QY      3 HPOFEKXXXXXXXXXWMSHPQFEK 25
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
DB      183 YSDFSEAACKVNNNSWMSHPQFEK 205

RESULT 7
AAB46421 AAB46421 standard; protein; 254 AA.
XX
AC AAB46421;
DT 06-APR-2001 (first entry)
DE Bilin binding-protein associated protein #2.
EE
EX Bilin-binding protein; muten; BBP; digoxigenin.
FY Unidentified.
GZ
ID WO200075308-A1.
IN
PI 14-DEC-2000.
PN
PP 08-JUN-2000; 2000WO-DE001873.
PR 08-JUN-1999; 99DE-01026068.
PS (SKER/) SKERRA A.
PT Skerra A, Schlenhuber S;
PX WPI: 2001-071071/08.
PY N-PSDB; AAF25703.
ZZ

PT New mutens of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
PS Example 1; Page 59-60; 80pp; German.
XX
XX This invention describes novel polypeptides (I) that are mutens of bilin
CC -binding protein (BBP), which can bind digoxigenin (dig) or its
CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
CC have an amino acid (aa) substitution at at least one of the positions 28,
CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
CC fusion proteins, are used to bind, detect, determine, immobilize or
CC separate Dig or its conjugates with proteins, nucleic acids,
CC carbohydrates, other biological or synthetic macromolecules or low
CC molecular weight compounds, particularly in assays where Dig is being
CC used as a label. Compared with Dig-specific antibodies, (I) have a
CC simpler structure and are easier to prepare. They have very high
CC specificity for Dig, relative to other steroids, and fusion partners may
CC be attached to either end without compromising their ability to bind
CC ligand
XX
SQ      Sequence 254 AA;                               61.3%; Score 68; DB 4; Length 254;
Query Match Similarity 39.1%; Pred. No. 2.9;
Best Local Similarity 39.1%; Pred. No. 2.9;
Matches          9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

QY      3 HPOFEKXXXXXXXXXWMSHPQFEK 25
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
DB      183 YSDFSEAACKVNNNSWMSHPQFEK 205

RESULT 8
```

XX	AAW93966		
ID	AAW93966 standard; protein; 396 AA.		
AC			
XX	AAW93966;		
DT			
XX	02-JUL-1999 (first entry)		
DE	Plasmid pBBP20 protein fragment.		
KW	Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;		
XV	tissue-specific surface molecule; tumour-specific surface molecule;		
XX	tumour imaging; bilin binding protein; phage coat protein pIII.		
OS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..21	
FT		/label= signal_peptide	
FT	Protein	22..396	
FT		/label= fusion_peptide	
FT		/note= "This fusion peptide is constructed from bilin	
FT		binding protein, a strep-Tag II region and a fragment of	
FT		phage coat protein pIII"	
FT	Protein	22..195	
FT		/note= "Bilin binding protein"	
FT	Region	196..205	
FT		/note= "Strep-Tag II region"	
FT	Misc-difference	206	
FT		/note= "Encoded by TAG"	
FT	Protein	207..396	
FT		/note= "phage coat protein pIII"	
XX			
PN	WO9916873-A1.		
PD	08-APR-1999.		
XX			
PF	25-SEP-1998; 98MO-DE002898.		
XX			
PR	26-SEP-1997; 97DE-01042706.		
PA	(SKERR/) SKERRA A.		
PI	Skerra A, Beese G, Schmidt F, Scibora T;		
DR	WPI; 1999-295093/21.		
DR	N-PDB; AAX24099.		
XX			
PT	New anticalins produced by mutation of lipocalin family protein.		
XX			
PS	Disclosure; Page 67-69; 80pp; German.		
CC	This invention describes the preparation of anticalins from polypeptides		
CC	of the lipocalin family by mutation of amino acids in the region of the 4		
CC	peptide loops (at one end of the cylindrical 'leaflet' structure).		
CC	Anticalins can bind to a predetermined ligand with determinable affinity.		
CC	Anticalins or fusions of anticalins, are used: (a) when immobilized, for		
CC	separation; and (b) when labeled, for identification of ligands (or their		
CC	fusions or conjugates). Typically they bind to tissue- or tumor-specific		
CC	surface molecules and can be used for tumour imaging or directly for		
CC	tumour therapy. Mutations can be introduced into lipocalin polypeptides		
CC	more easily than into antibodies, since lipocalin polypeptides have only		
CC	4 (contrast 6) sequence segments. The peptide loops can tolerate amino		
CC	acid changes without significant effect on folding. Anticalins have high		
CC	specific affinity, comparable with that for antibodies. This sequence		
CC	represents a fusion polypeptide used in the method of the invention		
XX			
SO	Sequence 396 AA;		
Query Match	61.3%; Score 68; DB 2; Length 396;		
Best Local Similarity	39.1%; Pred. No. 5.3;		
Matches	9; Conservative 11; Mismatches 3; Indels 0; Gaps 0		
QY	3 HPQFEKXXXXXXXXXSHPOFER 25		

```

Db      : | ::::::::::::::| |||||
         183 YSDPSEAAACKVNNNSMWSHPQPEK 205

RESULT 9
ID AAB46420 standard; protein; 396 AA.
AC AAB46420;
DT 06-APR-2001 (first entry)
DE Bilin binding-protein associated protein.
KW Bilin-binding protein; muten; BBP; digoxigenin.
OS Unidentified.
PN WO200075308-A1.
PD 14-DEC-2000.
PF 08-JUN-2000; 2000WO-DE001873.
PR 08-JUN-1999; 99DE-01026068.
PX (SKER/) SKERRA A.
PY Skerra A, Schlehuber S;
DR WPI: 2001-071071/08.
N-PSDB; AAFZ5695.

New muteins of bilin-binding protein, useful for detecting digoxigenin
being used as label in e.g. binding assays, are very selective for
digoxigenin.

Example 1; Page 55-56; 80pp; German.

This invention describes novel polypeptides (I) that are muteins of bilin
-binding protein (BBP), which can bind digoxigenin (Dig) or its
conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
have an amino acid (aa) substitution at at least one of the positions 28,
31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
fusion proteins, are used to bind, detect, determine, immobilize or
separate Dig or its conjugates with proteins, nucleic acids,
carbohydrates, other biological or synthetic macromolecules or low
molecular weight compounds, particularly in assays where Dig is being
used as a label. Compared with Dig-specific antibodies, (I) have a
similar structure and are easier to prepare. They have very high
specificity for Dig, relative to other steroids, and fusion partners may
be attached to either end without compromising their ability to bind
ligand

SQ Sequence 396 AA;
Query Match          61.3%; Score 68; DB 4; Length 396;
Best Local Similarity 39.1%; Pred.No. 5.3;
Matches   9; Conservative 11; Mismatches    3; Indels    0; Gaps    0;

OY      3 HPOFEKXXXXXXXXXXSMGHPQPEK 25
       : | ::::::::::::::| |||||
       : | ::::::::::::::| |||||
DB      183 YSDPSEAAACKVNNNSMWSHPQPEK 205

RESULT 10
ID AAB46423 standard; protein; 396 AA.
AC AAB46423;
DT 06-APR-2001 (first entry)
```

```

DE      Bilin binding-protein associated protein #4.
XX
XX      Bilin-binding protein; mutetin; BBP; digoxigenin.
XX
OS      Unidentified.
XX
XX      WO200075308-A1.
XX
XX      14-DEC-2000.
XX
XX      08-JUN-2000; 2000WO-DE001873.
XX
XX      08-JUN-1999; 99DE-01026068.
XX
XX      (SKER/) SKERRA A.
XX
XX      Skerra A, Schlehuber S;
XX
XX      WPI, 2001-071071/08.
XX
XX      N-PSDB; AAF25708.
XX
XX      New mutetins of bilin-binding protein, useful for detecting digoxigenin
XX      being used as label in e.g. binding assays, are very selective for
XX      digoxigenin.
XX
XX      Example 2; Page 62-64; 80pp; German.
XX
XX      This invention describes novel polypeptides (I) that are mutetins of bilin
XX      -binding protein (BBP), which can bind digoxigenin (Dig) or its
XX      conjugates, do not bind cubabin, testosterone or 4-aminofluorescein, and
XX      have an amino acid (aa) substitution at at least one of the positions 28,
XX      31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
XX      fusion proteins, are used to bind, detect, determine, immobilize or
XX      separate Dig or its conjugates with proteins, nucleic acids,
XX      carbohydrates, other biological or synthetic macromolecules or low
XX      molecular weight compounds, particularly in assays where Dig is being
XX      used as a label. Compared with Dig-specific antibodies, (I) have a
XX      simpler structure and are easier to prepare. They have very high
XX      specificity for Dig, relative to other steroids, and fusion partners may
XX      be attached to either end without compromising their ability to bind
XX      ligand
XX
XX      Sequence 396 AA;
XX
XX      Query Match      61.3%; Score 68; DB 4; Length 396;
XX      Best Local Similarity 39.1%; Pred. No. 5.3;
XX      Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;
XX
XX      QY      3 HPOFEKXXXXXXXXXWMSHPQFEK 25
XX      : | : : : : : : : : : : : : : : : :
XX      183 YSDPSEAAACKVNNMNSHPQFEK 205

RESULT 11
AAB46425      AAB46425 standard; protein; 448 AA.
XX
XX      AAB46425;
XX
XX      06-APR-2001 (first entry)
XX
XX      Bilin binding-protein associated protein #6.
XX
XX      Bilin-binding protein; mutetin; BBP; digoxigenin.
XX
XX      Unidentified.
XX
XX      WO200075308-A1.
XX
XX      14-DEC-2000.
XX
XX      08-JUN-2000; 2000WO-DE001873.

```

```

PR      08-JUN-1999; 99DE-01026068.
XX
XX      (SKER/) SKERRA A.
XX
XX      Skerra A, Schlehuber S;
XX
XX      WPI, 2001-071071/08.
XX
XX      N-PSDB; AAF25710.
XX
XX      New mutetins of bilin-binding protein, useful for detecting digoxigenin
XX      being used as label in e.g. binding assays, are very selective for
XX      digoxigenin.
XX
XX      Example 3; Page 65-67; 80pp; German.
XX
XX      This invention describes novel polypeptides (I) that are mutetins of bilin
XX      -binding protein (BBP), which can bind digoxigenin (Dig) or its
XX      conjugates, do not bind cubabin, testosterone or 4-aminofluorescein, and
XX      have an amino acid (aa) substitution at at least one of the positions 28,
XX      31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
XX      fusion proteins, are used to bind, detect, determine, immobilize or
XX      separate Dig or its conjugates with proteins, nucleic acids,
XX      carbohydrates, other biological or synthetic macromolecules or low
XX      molecular weight compounds, particularly in assays where Dig is being
XX      used as a label. Compared with Dig-specific antibodies, (I) have a
XX      simpler structure and are easier to prepare. They have very high
XX      specificity for Dig, relative to other steroids, and fusion partners may
XX      be attached to either end without compromising their ability to bind
XX      ligand
XX
XX      Sequence 448 AA;
XX
XX      Query Match      61.3%; Score 68; DB 4; Length 448;
XX      Best Local Similarity 39.1%; Pred. No. 6.3;
XX      Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;
XX
XX      QY      3 HPOFEKXXXXXXXXXWMSHPQFEK 25
XX      : | : : : : : : : : : : : : : : : :
XX      183 YSDPSEAAACKVNNMNSHPQFEK 205

RESULT 12
AAB46426      AAB46426 standard; protein; 659 AA.
XX
XX      AAB46426;
XX
XX      06-APR-2001 (first entry)
XX
XX      Bilin binding-protein associated protein #7.
XX
XX      Bilin-binding protein; mutetin; BBP; digoxigenin.
XX
XX      Unidentified.
XX
XX      WO200075308-A1.
XX
XX      14-DEC-2000.
XX
XX      08-JUN-2000; 2000WO-DE001873.
XX
XX      08-JUN-1999; 99DE-01026068.
XX
XX      (SKER/) SKERRA A.
XX
XX      Skerra A, Schlehuber S;
XX
XX      WPI, 2001-071071/08.
XX
XX      N-PSDB; AAF25711.
XX
XX      New mutetins of bilin-binding protein, useful for detecting digoxigenin
XX      being used as label in e.g. binding assays, are very selective for
XX      digoxigenin.

```

XX Example 4; Page 68-71; 80pp; German.

PS This invention describes novel polypeptides (I) that are muteins of bilin

CC -binding protein (BBP), which can bind digoxigenin (Dig) or its

CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and

CC have an amino acid (aa) substitution at at least one of the positions 28,

CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its

CC fusion proteins, are used to bind, detect, determine, immobilize or

CC separate Dig or its conjugates with proteins, nucleic acids,

CC carbohydrates, other biological or synthetic macromolecules or low

CC molecular weight compounds, particularly in assays where Dig is being

CC used as a label. Compared with Dig-specific antibodies, (I) have a

CC simpler structure and are easier to prepare. They have very high

CC specificity for Dig, relative to other steroids, and fusion partners may

CC be attached to either end without compromising their ability to bind

CC ligand

XX Sequence 659 AA;

SQ

Query Match 61.3%; Score 68; DB 4; Length 659;
Best Local Similarity 39.1%; Pred.No. 11;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

OY 3 HPOEFKXXXXXXXXXWHPQFEK 25
 : | :::::| |||||
DB 637 YSDFSEACKVNNMNSHPOFEK 659

RESULT 13

AAB46427 AAB46427 standard; protein; 659 AA.
XX AC AAB46427;
XX DT 06-APR-2001 (first entry)
XX DE Bilin binding-protein associated protein #8.
XX KW Bilin-binding protein; mutenin; BBP; digoxigenin.
XX OS Unidentified.
XX PN WO300075308-A1.
XX PD 14-DEC-2000.
XX PF 08-JUN-2000; 2000WO-DE001873.
XX PR 08-JUN-1999; 99DE-01026068.
XX PA (SKER/) SKERRA A.
PI Skerra A, Schlehuber S;
DR MPI; 2001-071071/08.
DR N-Psdb; AAF25712.
XX PT New muteins of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin

CC -binding protein (BBP), which can bind digoxigenin (Dig) or its

CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and

CC have an amino acid (aa) substitution at at least one of the positions 28,

CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its

CC fusion proteins, are used to bind, detect, determine, immobilize or

CC separate Dig or its conjugates with proteins, nucleic acids,

CC carbohydrates, other biological or synthetic macromolecules or low

CC molecular weight compounds, particularly in assays where Dig is being

CC used as a label. Compared with Dig-specific antibodies, (I) have a

CC simpler structure and are easier to prepare. They have very high

CC specificity for Dig, relative to other steroids, and fusion partners may

CC be attached to either end without compromising their ability to bind

CC ligand

XX Sequence 659 AA;

SQ

Query Match 61.3%; Score 68; DB 4; Length 659;
Best Local Similarity 39.1%; Pred.No. 11;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

OY 3 HPOEFKXXXXXXXXXWHPQFEK 25
 : | :::::| |||||
DB 637 YSDFSEACKVNNMNSHPOFEK 659

RESULT 13

AAB46427 AAB46427 standard; protein; 659 AA.
XX AC AAB46427;
XX DT 06-APR-2001 (first entry)
XX DE Bilin binding-protein associated protein #8.
XX KW Bilin-binding protein; mutenin; BBP; digoxigenin.
XX OS Unidentified.
XX PN WO300075308-A1.
XX PD 14-DEC-2000.
XX PF 08-JUN-2000; 2000WO-DE001873.
XX PR 08-JUN-1999; 99DE-01026068.
XX PA (SKER/) SKERRA A.
PI Skerra A, Schlehuber S;
DR MPI; 2001-071071/08.
DR N-Psdb; AAF25712.
XX PT New muteins of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin

CC -binding protein (BBP), which can bind digoxigenin (Dig) or its

CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and

CC have an amino acid (aa) substitution at at least one of the positions 28,

CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its

CC fusion proteins, are used to bind, detect, determine, immobilize or

CC separate Dig or its conjugates with proteins, nucleic acids,

CC carbohydrates, other biological or synthetic macromolecules or low

CC molecular weight compounds, particularly in assays where Dig is being

CC used as a label. Compared with Dig-specific antibodies, (I) have a

CC simpler structure and are easier to prepare. They have very high

CC specificity for Dig, relative to other steroids, and fusion partners may

CC be attached to either end without compromising their ability to bind

CC ligand

XX Sequence 659 AA;

SQ

Query Match 61.3%; Score 68; DB 4; Length 659;
Best Local Similarity 39.1%; Pred.No. 11;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

OY 3 HPOEFKXXXXXXXXXWHPQFEK 25
 : | :::::~| |||||
DB 637 YSDFSEACKVNNMNSHPOFEK 659

RESULT 13

AAB46427 AAB46427 standard; protein; 659 AA.
XX AC AAB46427;
XX DT 06-APR-2001 (first entry)
XX DE Bilin binding-protein associated protein #8.
XX KW Bilin-binding protein; mutenin; BBP; digoxigenin.
XX OS Unidentified.
XX PN WO300075308-A1.
XX PD 14-DEC-2000.
XX PF 08-JUN-2000; 2000WO-DE001873.
XX PR 08-JUN-1999; 99DE-01026068.
XX PA (SKER/) SKERRA A.
PI Skerra A, Schlehuber S;
DR MPI; 2001-071071/08.
DR N-Psdb; AAF25712.
XX PT New muteins of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin

CC -binding protein (BBP), which can bind digoxigenin (Dig) or its

CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and

CC have an amino acid (aa) substitution at at least one of the positions 28,

CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its

CC fusion proteins, are used to bind, detect, determine, immobilize or

CC separate Dig or its conjugates with proteins, nucleic acids,

CC carbohydrates, other biological or synthetic macromolecules or low

CC molecular weight compounds, particularly in assays where Dig is being

CC used as a label. Compared with Dig-specific antibodies, (I) have a

CC simpler structure and are easier to prepare. They have very high

CC specificity for Dig, relative to other steroids, and fusion partners may

CC be attached to either end without compromising their ability to bind

CC ligand

XX Sequence 659 AA;

SQ

Query Match 61.3%; Score 68; DB 4; Length 659;
Best Local Similarity 39.1%; Pred.No. 11;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

OY 3 HPOEFKXXXXXXXXXWHPQFEK 25
 : | :::::~| |||||
DB 637 YSDFSEACKVNNMNSHPOFEK 659

RESULT 13

AAB46427 AAB46427 standard; protein; 659 AA.
XX AC AAB46427;
XX DT 06-APR-2001 (first entry)
XX DE Bilin binding-protein associated protein #8.
XX KW Bilin-binding protein; mutenin; BBP; digoxigenin.
XX OS Unidentified.
XX PN WO300075308-A1.
XX PD 14-DEC-2000.
XX PF 08-JUN-2000; 2000WO-DE001873.
XX PR 08-JUN-1999; 99DE-01026068.
XX PA (SKER/) SKERRA A.
PI Skerra A, Schlehuber S;
DR MPI; 2001-071071/08.
DR N-Psdb; AAF25712.
XX PT New muteins of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin

CC -binding protein (BBP), which can bind digoxigenin (Dig) or its

CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and

CC have an amino acid (aa) substitution at at least one of the positions 28,

CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its

CC fusion proteins, are used to bind, detect, determine, immobilize or

CC separate Dig or its conjugates with proteins, nucleic acids,

CC carbohydrates, other biological or synthetic macromolecules or low

CC molecular weight compounds, particularly in assays where Dig is being

CC used as a label. Compared with Dig-specific antibodies, (I) have a

CC simpler structure and are easier to prepare. They have very high

CC specificity for Dig, relative to other steroids, and fusion partners may

CC be attached to either end without compromising their ability to bind

CC ligand

XX Sequence 659 AA;

SQ

Query Match 61.3%; Score 68; DB 4; Length 659;
Best Local Similarity 39.1%; Pred.No. 11;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

OY 3 HPOEFKXXXXXXXXXWHPQFEK 25
 : | :::::~| |||||
DB 637 YSDFSEACKVNNMNSHPOFEK 659

RESULT 13

AAB46427 AAB46427 standard; protein; 659 AA.
XX AC AAB46427;
XX DT 06-APR-2001 (first entry)
XX DE Bilin binding-protein associated protein #8.
XX KW Bilin-binding protein; mutenin; BBP; digoxigenin.
XX OS Unidentified.
XX PN WO300075308-A1.
XX PD 14-DEC-2000.
XX PF 08-JUN-2000; 2000WO-DE001873.
XX PR 08-JUN-1999; 99DE-01026068.
XX PA (SKER/) SKERRA A.
PI Skerra A, Schlehuber S;
DR MPI; 2001-071071/08.
DR N-Psdb; AAF25712.
XX PT New muteins of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin

CC -binding protein (

```

CC used as a label. Compared with Dig-specific antibodies, (I) have a
CC simpler structure and are easier to prepare. They have very high
CC specificity for Dig, relative to other steroids, and fusion partners may
CC be attached to either end without compromising their ability to bind
CC ligand
CC
SQ Sequence 659 AA:
XX
XX
XX Query Match      61.3%, Score 68; DB 4; Length 659;
   Best Local Similarity    39.1%, Pred.No. 11;
   Matches    9; Conservative    11; Mismatches    3; Indels    0; Gaps    0;
OY      3 HPQFEKXXXXXXXXXMSHPQFEK 25
       :|::|::::|::|::|::|::|
DB      183 YSDFSEAAACKVNNSMWSHPQFEK 205
RESULT 14
ADB87j310
ID ADB87j310 standard; protein; 248 AA.
XX
XX ADB87j310;
AC
XX 04-DEC-2003 (first entry)
DT
DE Apolipoprotein D/Strep-Tag II/albumin binding domain fusion protein.
XX
XX Human; apolipoprotein D; Apod; lipocalin; serum lipid transport;
KW albumin binding domain; Protein G.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH 1..21
FT Peptide /note= "Signal peptide"
FT 22..248
FT Protein /note= "Apolipoprotein D/Strep-Tag II/albumin binding
FT domain fusion protein"
FT 22..190
FT Protein /note= "Mature Apod"
FT 191..199
FT Peptide /note= "Strep-Tag II affinity tag"
FT 200..248
FT Protein /note= "Albumin binding domain of Protein G"
PN WO2003029471-A1.
XX
XX 10-APR-2003.
PD
XX 27-SEP-2001; 2001WO-EP011211.
PF
XX 27-SEP-2001; 2001WO-EP011211.
PR
XX (PIER-) PIERIS PROTEOLAB AG.
PA
XX
XX Vogt M, Skerra A;
PI
XX WPI; 2003-381641/36.
DR N-Psdb; ADB87298.
DR
XX Generating a mutein of human apolipoprotein D having detectable affinity
PT to a given non-natural ligand of apolipoprotein D by subjecting
PT apolipoprotein D to mutagenesis.
XX
XX Disclosure: Page 64-65; 70pp; English.
PS
XX The present invention relates to a human apolipoprotein D (Apod) mutein
CC (ADB87307) having detectable affinity to a given non-natural ligand of
CC Apod. Apod is a member of the lipocalin family of proteins, and is
CC involved in serum lipid transport. Mutein ADB87307 can be used as a
CC pharmacological drug target. The present sequence, encoded by a fragment
CC of plasmid pApod12, is a fusion protein of Apod, a Strep-Tag II and the
CC albumin binding domain of Protein G.
```

xx	Sequence	248	AA;
sq	Query Match	60.4%;	Score 67; DB 7; Length 248;
	Best Local Similarity	36.0%;	Pred. No. 3.9;
	Matches	9; Conservative	13; Mismatches 3; Indels 0; Gaps 0;
Qy	1	WSHPQFEKXXXXXXXXXXWSHPQFEK	25
Db	192	WSHPQFEKPSLAENKVLNRELDK	216

RESULT 15	
ADA27293	
ID	ADA27293 standard; protein; 258 AA.
XX	

DT 20-NOV-2003 (first entry)
XX

Plasmid pHNGAL7 fragment protein.

alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.

OS Synthetic.

	Key	Location/Qualifiers
FH	Pentide	1 31
ET		

```

ET      /note= "Ompa signal sequence"
ET      33 373
Protein

```

/note= "Mature protein; fusion protein of hNGAL, Strep-Tag II and albumin binding domain"

FT	Protein	22.	.199
EM			

FT	Peptide	/note="Mature hNGAL"
199	199	199

[illegible]

PN WO2003029462-A1

PD 10-APR-2003.

27-SEP-2001; 2001WO-EP011213.

27-SEP-2001; 2001WO-EP011213.

(PIER-) PIERIS PROTEOLAB AG.

Skerra A, Schlehuber S;

WPI; 2003-381639/36.
N-Depend: 40307007

PT XX

targeted by subjecting the protein to mutagenesis at sequence positions corresponding to sequence positions of the human neutrophil gelatinase associated lipocalin.

Disclosure; Page 59-61; 68pp; English.

The present invention relates to a method for generating mutants of human neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-microglobulin-related protein (A α m) or mouse 24p3/ucrocalin, by subjecting the protein to mutagenesis. In the method of generating a mutant, a nucleic acid coding for the mutants of the protein resulting from mutagenesis is operably fused at the 3' end with a gene coding for the coat protein of a filamentous bacteriophage of the M13-family or for a fragment of the coat protein. The present sequence is the protein encoded by a fragment of plasmid pHNGAL¹, used to illustrate the invention. This sequence comprises human hNGAL, a Strep-Tag II and an albumin binding domain of protein G.

XX	Sequence	258 AA;
SQ	Query Match	60.4%; Score 67; DB 6; Length 258; Beet Local Similarity 36.0%; Pred. No. 4.1; Matches 9; Conservative 13; Mismatches 3; Indels 0; Gaps 0;
OY	1 WSHPOFEKXXXXXXXXXWSHPQPEK 25 :::.. Db 202 WSHPOFEKPASLAKVLANREDUK 226	

```
Search completed: March 2, 2005, 13:02:49
Job time : 113.602 secs
```

Page 8

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 77.7439 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ13
Perfect score: 111
Sequence: 1 wshpqfexkxxxxxxwshpqfex 25

Scoring table: BLOSUM62GX
Gapop 10.0, Gapext 0.5

Searched: 1385339 seqs, 328044528 residues
Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	89.6	24	US-10-026-578B-3	Sequence 3, Appl1
2	95.5	86.0	36	US-10-026-578B-10	Sequence 10, Appl1
3	82.5	74.3	36	US-10-026-578B-11	Sequence 11, Appl1
4	71	64.0	845	US-10-628-432-40	Sequence 40, Appl1
5	68	61.3	21	US-09-809-517A-30	Sequence 30, Appl1
6	68	61.3	22	US-09-809-517A-33	Sequence 33, Appl1
7	64.5	58.1	646	US-10-628-432-49	Sequence 49, Appl1
8	64.5	58.1	858	US-10-628-432-27	Sequence 27, Appl1
9	63	56.8	485	US-10-628-432-47	Sequence 47, Appl1
10	63	56.8	663	US-10-344-741-2	Sequence 2, Appl1
11	63	56.8	697	US-10-628-432-24	Sequence 24, Appl1
12	62.5	56.3	24	US-09-809-517A-31	Sequence 31, Appl1
13	62.5	56.3	25	US-09-809-517A-34	Sequence 34, Appl1

14	62.5	56.3	245	US-10-887-228A-1	Sequence 1, Appl1
15	62.5	56.3	246	US-10-887-228A-9	Sequence 9, Appl1
16	62.5	56.3	252	US-10-887-228A-5	Sequence 5, Appl1
17	62	55.9	117	US-09-977-137A-4	Sequence 4, Appl1
18	62	55.9	117	US-09-977-137A-5	Sequence 5, Appl1
19	62	55.9	117	US-09-977-137A-7	Sequence 7, Appl1
20	62	55.9	117	US-09-977-137A-8	Sequence 8, Appl1
21	62	55.9	117	US-09-977-137A-9	Sequence 9, Appl1
22	62	55.9	117	US-09-977-137A-10	Sequence 10, Appl1
23	62	55.9	117	US-09-977-137A-11	Sequence 11, Appl1
24	62	55.9	118	US-09-977-137A-12	Sequence 12, Appl1
25	62	55.9	118	US-09-977-137A-6	Sequence 6, Appl1
26	61	55.0	633	US-10-628-432-53	Sequence 53, Appl1
27	60.5	54.5	261	US-10-767-701-42119	Sequence 42119, A
28	60	54.1	24	US-10-026-578B-4	Sequence 4, Appl1
29	60	54.1	661	US-10-354-983-8	Sequence 8, Appl1
30	60	54.1	763	US-10-358-283-15	Sequence 15, Appl1
31	57.5	51.8	86	US-09-884-767A-9	Sequence 9, Appl1
32	56	50.5	87	US-10-424-599-272001	Sequence 272001, A
33	55	49.5	112	US-10-424-599-171374	Sequence 171374, A
34	55	49.5	126	US-10-437-963-142983	Sequence 142983, A
35	55	49.5	134	US-10-437-963-104065	Sequence 104065, A
36	55	49.5	158	US-10-424-599-184769	Sequence 184769, A
37	55	49.5	212	US-10-424-599-184075	Sequence 184075, A
38	55	49.5	252	US-10-437-963-200150	Sequence 200150, A
39	55	49.5	315	US-10-369-493-5628	Sequence 5628, A
40	55	49.5	318	US-10-219-810-46	Sequence 46, Appl1
41	55	49.5	318	US-10-380-132-22	Sequence 22, Appl1
42	55	49.5	318	US-10-471-243-20	Sequence 20, Appl1
43	55	49.5	330	US-10-219-810-47	Sequence 47, Appl1
44	54.5	49.1	143	US-10-437-963-196479	Sequence 196479, A
45	54.5	49.1	381	US-10-425-114-63501	Sequence 63501, A

ALIGNMENTS

RESULT 1
US-10-026-578B-3
Sequence 3, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Traps
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (10)..(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:

```

? NAME/KEY: MISC_FEATURE
? LOCATION: (12)..(12)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (13)..(13)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (14)..(14)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (15)..(15)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (16)..(16)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Artificial Sequence represents peptide binding module
?S-10-026-578B-3

```

Query Match	89.6%	Score 99.5;	DB 14;	Length 24;
Best Local Similarity	96.0%;	Pred. No. 2.1e-05;		
Matches 24;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1

OY 1 WSHPQFEKXXXXXXXXXXWSHPQFEK 25
||| ||| ||| ||| ||| ||| ||| |||
Db 1 WSHPQFEK-XXXXXXXXXXWSHPQFEK 24

RESULT 2
US-10-026-578B-10
; Sequence 10, Application US/10026578B
; Publication No. US20030083474A1

? APPLICANT: Schmidt, Thomas
 ? TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tags
 ? FILE REFERENCE: 100810.01US1
 ? CURRENT APPLICATION NUMBER: US/10/026,579B
 ? PRIOR FILING DATE: 2002-11-11
 ? PRIOR APPLICATION NUMBER: DE 101 13 776.1
 ? PRIOR FILING DATE: 2001-03-21
 ? PRIOR APPLICATION NUMBER: PCT/EP01/11846
 ? PRIOR FILING DATE: 2001-10-12
 ? NUMBER OF SEQ ID NOS: 14
 ? SOFTWARE: PatentIn version 3.1

```

1 ORGANISM: Artificial Sequence
2 FEATURE:
3 / OTHER INFORMATION: Synthetic peptide
4 FEATURE:
5 / NAME/KEY: MISC FEATURE
6 / LOCATION: (9)..(28)
7 / OTHER INFORMATION: x represents a single amino acid at each of the positions indicated
8 / OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are
9 / OTHER INFORMATION: missing, the total numbers of x will be no less than 5
10 FEATURE:
11 / NAME/KEY: misc_feature
12 / OTHER INFORMATION: Artificial Sequence represents peptide binding module
13 / OS:10-026-578B-10

```

```

Query Match      86.0%; Score 95.5; DB 14; Length 36;
Best Local Similarity 69.4%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 11; Gaps 1
QY 1 WSHPQPEK-----XXXXXXXXXXWSHPQPEK 25
|||||

```

RESULT 3
US-10-026-578B-11
; Sequence 11, Application US/10026578B
; Publication No. US20030083474A1

```

1 TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
2 FILE REFERENCE: 100810.01u51
3 CURRENT APPLICATION NUMBER: US/10/026,578B
4 CURRENT FILING DATE: 2002-11-11
5 PRIOR APPLICATION NUMBER: DE 101 13 776.1
6 PRIOR FILING DATE: 2001-03-21
7 PRIOR APPLICATION NUMBER: PCT/EP01/11846
8 PRIOR FILING DATE: 2001-10-12
9 NUMBER OF SEQ ID NOS: 14
10 SOFTWARE: PatentIn version 3.1
11 SEQ ID NO: 11

```

```

1 ORGANISM: Artificial Sequence
2 FEATURE:
3 OTHER INFORMATION: Synthetic Peptide
4 FEATURE:
5 NAME/KEY: MISC_FEATURE
6 LOCATION: (9)..(28)
7 OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
8 OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
9 OTHER INFORMATION: be present

```

OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11

Query Match	74.3%	Score 82.5;	DB 14;	Length 36;
Best Local Similarity	41.7%;	Pred. No. 0.0053;		
Matches 15;	Conservative 9;	Mismatches 1;	Indels 11;	Gaps 1;

```
QY 1 WSHQPFKEK-----XXXXXXXXXXWSHPQFEK 25
      |||||
      ::::: |||||
Db 1 WSHQPFKEKGGSGSGSGSGSGSGSGSQSHQPFKEK 36
```

RESULT 4
US-10-628-432-40
; Sequence 40, Application US/10628432
; Publication No. US20040142863A1

```

1 TITLE OF INVENTION: Modified ADAMTS4 molecules
2
3 FILE REFERENCE: AM101378
4
5 CURRENT APPLICATION NUMBER: US/10/628,432
6
7 CURRENT FILING DATE: 2003-07-29
8
9 NUMBER OF SEQ ID NOS: 53
10
11 SOFTWARE: PatentIn version 3.1
12
13 SEQ ID NO: 40
14
15 LENGTH: 845
16
17 TYPE: PRT
18
19 ORGANISM: Artificial
20
21 FEATURES:
22
23 OTHER INFORMATION: ADAMTS4 ASM with insertion
24
25 US-10-628-432-40

```

Query Match	64.0%;	Score 71;	DB 16;	Length 845;
Best Local Similarity	50.0%;	Pred. No. 7;		
Matches	9;	Mismatches	0;	Gaps 0;
QY	1 WSHPOFEKXXXXXXXXXW	18		
b				
520	WSHPOFEKAGWGPMPM	537		
	:::::::			

```

RESULT 5
US-09-809-517A-30
Sequence 30. US2002003473A1
Patent No. US2002003473A1
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. US2002003473A1 methods for displaying (poly)peptides/proteins
TITLE OF INVENTION: particles via disulfide bonds
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: Patentin version 3.0
SEQ ID NO 30
LENGTH: 21
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

```

	Query Match	61.3%	Score 68;	DB 9;	Length 21;
	Best Local Similarity	42.9%	Pred. No. 0.2;		
Matches	9;	Conservative	11;	Mismatches	1;
				Indels	0;
Gy	5 QFEKXXXXXXXXXSHPOFEK	25			
	:::::				
Dh	1 EFDDYKDDDKGAPWSHPOFEK	21			

```

RESULT 6
US-09-809-517A-33
Sequence 33, Application US/09809517A
Patent No. US2002003473A1
GENERAL INFORMATION:
APPLICANT: Lohning, Corina
TITLE OF INVENTION: No. US2002003473A1 method for displaying (poly) peptides/prote
TITLE OF INVENTION: particles via disulfide bonds
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 33
LENGTH: 22
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

```

	Query Match	61.3%	Score 68	DB 9	Length 22
	Best Local Similarity	42.9%	Pred No.	0.21	
	Matches	9	Conservative	11	Mismatches 1; Indels 0; Gaps 0
Oy	5 QPEKXXXXXXXXXXSHPOPEK	25			
	: : : : : : : : : : : : : :				
Db	2 EFDYDDDDDKGAPWMSHPQPEK	22			

RESULT 7
US-10-628-432-49
; Sequence 49, Application US/10628432

```

? Publication No. US20040142863A1
? GENERAL INFORMATION:
? APPLICANT: Myeth
? TITLE OF INVENTION: Modified ADAMTS4 molecules
? FILE REFERENCE: AM101378
? CURRENT APPLICATION NUMBER: US/10/628,432
? CURRENT FILING DATE: 2003-07-29
? NUMBER OF SEQ ID NOS: 53
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 49
? LENGTH: 646
? TYPE: PRT
? ORGANISM: Artificial
? FEATURE:
? OTHER INFORMATION: furin-processed construct E
US-10-628-432-49

Query Match          59.1%; Score 64.5; DB 16; Length 646;
Best Local Similarity 31.2%; Pred. No. 35;
Matches 10; Conservative 12; Mismatches 3; Indels 7; Gaps 1;

```

```

Qy      1 WSHPQFE-----KXXXXXXXXXXWSHPQFEK 25
        | : : :::::|||||
Db      615 WLHRRAQLEILRRPFWAGKGSAMSHPQFEK 646

```

```

RESULT 8
US-10-628-432-27
: Sequence 27, Application US/10628432
: Publication No. US20040142863A1
: GENERAL INFORMATION:
: APPLICANT: Wyeth
: TITLE OF INVENTION: Modified ADAMTS4 molecules
: FILE REFERENCE: AM101378
: CURRENT APPLICATION NUMBER: US/10/628,432
: CURRENT FILING DATE: 2003-07-29
: NUMBER OF SEQ ID NOS: 53
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 27
: LENGTH: 858
: TYPE: PRT
: ORGANISM: Artificial
: FEATURE:
: OTHER INFORMATION: modified ADAMTS4 molecule
: US-10-628-432-27

```

Query Match	Similarity	58.1%	Score 64.5	DB 16	Length 858
Best Local	Similarity	31.2%	Pred. No. 49		
Matches	10	Conservative	12	Mismatches	3
				Indels	7
				Gaps	1
Qy	1	WSHPQFE-----XXXXXXXWSHPQFEK	25		
		: : : : :			
		: : : : :			
Db	827	WLHRRQILLETILRRPWPAGRGASWSHPQFEK	858		

```

RESULT 9
US-10-628-432-47
Sequence 47, Application US/10628432
Publication NO. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47
LENGTH: 485
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: furin-processed construct C

```

US-10-628-432-47

Query Match 56.8%; Score 63; DB 16; Length 485;
Best Local Similarity 36.7%; Pred. No. 38;
Matches 11; Conservative 11; Mismatches 2; Indels 6; Gaps 1;

OY 2 SHPOFEK-----XXXXXXXXXXSHPOFEK 25
DB 456 SKKFFDKMVCGGDGGSGSAMSHPOFEK 485

RESULT 10
US-10-344-741-2
; Sequence 2, Application US/10344741
; Publication No. US20040038371A1
; GENERAL INFORMATION:
; APPLICANT: Basten, Danielle
; APPLICANT: Dekker, Petrus, Jacobus, Theodorus
; APPLICANT: Schuurhuizen, Paul, William
; APPLICANT: Schaap, Petrus, Johannes
; APPLICANT: Visser, Jacob
; APPLICANT: DSM NV
; TITLE OF INVENTION: No. US20040038371A1 Amino-peptidase
; FILE REFERENCE: 24615-20187.00
; CURRENT APPLICATION NUMBER: US/10/344,741
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: EP/NL 00202995.7
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Aspergillus niger N400

US-10-344-741-2

Query Match 56.8%; Score 63; DB 15; Length 663;
Best Local Similarity 31.8%; Pred. No. 56;
Matches 7; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPO 22
DB 176 YSHPTFSPAGDRVSWIQNHPE 197

RESULT 11
US-10-628-432-24
; Sequence 24, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 24
; LENGTH: 697
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Truncated ADAMTS4 molecule
US-10-628-432-24

Query Match 56.8%; Score 63; DB 16; Length 697;
Best Local Similarity 36.7%; Pred. No. 59;
Matches 11; Conservative 11; Mismatches 2; Indels 6; Gaps 1;

OY 2 SHPOFEK-----XXXXXXXXXXSHPOFEK 25
DB 668 SKKFFDKMVCGGDGGSGSAMSHPOFEK 697

RESULT 12
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match 56.3%; Score 62.5; DB 9; Length 24;
Best Local Similarity 41.7%; Pred. No. 1.2;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

OY 5 OFEK-----XXXXXXXXXXSHPOFEK 25
DB 1 EFEOKLISEEDLNGAPWMSHPOFEK 24

RESULT 13
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match 56.3%; Score 62.5; DB 9; Length 25;
Best Local Similarity 41.7%; Pred. No. 1.3;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

OY 5 OFEK-----XXXXXXXXXXSHPOFEK 25
DB 2 EFEOKLISEEDLNGAPWMSHPOFEK 25

RESULT 14
US-10-887-228A-1
; Sequence 1, Application US/10887228A
; Publication No. US20050037402A1

```
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ TITLE OF INVENTION: Substance that is Contained in a Sample
/ FILE REFERENCE: S305690S
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ CURRENT FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ PRIOR FILING DATE: 2003-07-16
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 245
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-1

Query Match          56.3%; Score 62.5; DB 17; Length 245;
Best Local Similarity 41.7%; Pred.No. 20;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

QY 5 OPEK---XXXXXXXSHPOPEK 25
   :||: :||:|||||
Db 222 EFQKLISEEDLNGAPWSHPOPEK 245

RESULT 15
US-10-887-228A-9
/ Sequence 9, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ TITLE OF INVENTION: Substance that is Contained in a Sample
/ FILE REFERENCE: S305690S
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ CURRENT FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ PRIOR FILING DATE: 2003-07-16
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 9
/ LENGTH: 246
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match          56.3%; Score 62.5; DB 17; Length 246;
Best Local Similarity 41.7%; Pred.No. 20;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

QY 5 OPEK---XXXXXXXSHPOPEK 25
   :||: :||:|||||
Db 223 EFQKLISEEDLNGAPWSHPOPEK 246

Search completed: March 2, 2005, 14:19:02
Job time : 80.7439 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 28.4553 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ13

Perfect score: 111
Sequence: 1 wshpafekxxxxxxxxxwshpafek 25

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep: *
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep: *
5: /cgn2_6/prodata/1/1aa/5A_COMB.pep: *
6: /cgn2_6/prodata/1/1aa/5B_COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	61.3	21	US-09-809-517A-30	Sequence 30, Appl
2	68	61.3	22	US-09-809-517A-33	Sequence 33, Appl
3	62.5	56.3	24	US-09-809-517A-31	Sequence 31, Appl
4	62.5	56.3	25	US-09-809-517A-34	Sequence 34, Appl
5	62	55.9	117	US-09-809-517A-4	Sequence 4, Appl
6	62	55.9	117	US-09-809-517A-5	Sequence 5, Appl
7	62	55.9	117	US-09-809-517A-7	Sequence 7, Appl
8	62	55.9	117	US-09-809-517A-8	Sequence 8, Appl
9	62	55.9	117	US-09-809-517A-9	Sequence 9, Appl
10	62	55.9	117	US-09-809-517A-10	Sequence 10, Appl
11	62	55.9	117	US-09-809-517A-11	Sequence 11, Appl
12	62	55.9	117	US-09-809-517A-12	Sequence 12, Appl
13	62	55.9	117	US-09-809-517A-13	Sequence 13, Appl
14	55	48.6	155	US-09-809-517A-14	Sequence 14, Appl
15	55	48.6	155	US-09-809-517A-15	Sequence 15, Appl
16	54	48.6	155	US-09-809-517A-16	Sequence 16, Appl
17	54	48.6	155	US-09-809-517A-17	Sequence 17, Appl
18	54	48.6	155	US-09-809-517A-18	Sequence 18, Appl
19	54	48.6	155	US-09-809-517A-19	Sequence 19, Appl
20	54	48.6	155	US-09-809-517A-20	Sequence 20, Appl
21	54	48.6	155	US-09-809-517A-21	Sequence 21, Appl
22	54	48.6	155	US-09-809-517A-22	Sequence 22, Appl
23	54	48.6	155	US-09-809-517A-23	Sequence 23, Appl
24	54	48.6	155	US-09-809-517A-24	Sequence 24, Appl
25	54	48.6	155	US-09-809-517A-25	Sequence 25, Appl
26	54	48.6	155	US-09-809-517A-26	Sequence 26, Appl
27	54	48.6	155	US-09-809-517A-27	Sequence 27, Appl

28	52.5	47.3	699	4	US-09-538-092-995	Sequence 995, App
29	52	46.8	146	4	US-09-270-767-35770	Sequence 35770, A
30	52	46.8	146	4	US-09-270-767-35987	Sequence 35987, A
31	52	46.8	243	4	US-09-252-991A-26513	Sequence 26513, A
32	52	46.8	646	4	US-09-252-991A-26479	Sequence 26479, A
33	51	45.9	8	3	US-08-948-097-2	Sequence 2, Appl
34	51	45.9	8	3	US-09-382-950-7	Sequence 7, Appl
35	51	45.9	8	3	US-09-382-950-7	Sequence 8, Appl
36	51	45.9	8	4	US-09-619-103-9	Sequence 9, Appl
37	51	45.9	8	4	US-10-104-218-5	Sequence 5, Appl
38	51	45.9	8	4	US-09-809-517A-9	Sequence 9, Appl
39	51	45.9	155	4	US-08-887-534A-55	Sequence 55, Appl
40	51	45.9	155	4	US-09-527-431-55	Sequence 55, Appl
41	51	45.9	155	4	US-09-446-861-55	Sequence 55, Appl
42	51	45.9	604	3	US-09-586-935-3	Sequence 3, Appl
43	51	45.9	604	3	US-09-872-861-4	Sequence 4, Appl
44	51	45.9	754	4	US-09-198-452A-874	Sequence 874, App
45	51	45.9	754	4	US-09-438-185A-817	Sequence 817, App

ALIGNMENTS

```
RESULT 1
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PR
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match      61.3% Score 68; DB 4; Length 21;
Best Local Similarity 42.9%; Pred. No. 0.009;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;

QY      5 QPEKXXXXXXXXXWSPQFEK 25
       1 EFDYKDDDKGAPWSPQFEK 21
       1 EFDYKDDDKGAPWSPQFEK 21

RESULT 2
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

      / SOFTWARE: PatentIn version 3.0
      / SEQ ID NO 33
      / LENGTH: 22
      / TYPE: PRT
      / ORGANISM: artificial sequence
      / FEATURE:
      / OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match
Best Local Similarity    61.3%; Score 68; DB 4; Length 22;
Matches    9; Conservative   11; Mismatches    1; Indels    0; Gaps    0;

QY      5 QFEKXXXXXXXXXWSPQFEK 25
       :|:::|||||||
Db      2 EFDVKDDDKGAPWSHPQFEK 22

RESULT 3
US-09-809-517A-31
Sequence 31, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 34
LENGTH: 25
TYPE: PRT
ORGANISM: artificial sequence

US-09-809-517A-31
Description of Artificial Sequence: synthetic module

Query Match
Best Local Similarity    56.3%; Score 62.5; DB 4; Length 24;
Matches    10; Conservative   11; Mismatches    0; Indels    3; Gaps    1;

QY      5 QFER--XXXXXXXWSPQFEK 25
       :|:::|||||||
Db      1 EFEGKLISEEDLNGAPWSHPQFEK 24

RESULT 4
US-09-809-517A-34
Sequence 34, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 34
LENGTH: 25
TYPE: PRT
ORGANISM: artificial sequence

```

```

/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
- US-09-809-517A-34

Query Match          56.3%; Score 62.5; DB 4; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.072;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1

QY      5 OFEK---XXXXXXXWSHPQPEK 25
       :||: :|||||
Db      2 EFQKLISEEDLNGAPWSHPQPEK 25

RESULT 5
US-09-977-137A-4
/ Sequence 4, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Caguiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ TITLE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXWSHPQPEK 25
       :|||:|||||
Db      100 RKNVSCPSAMWSHPQPEK 117

RESULT 6
US-09-977-137A-5
/ Sequence 5, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Caguiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ TITLE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

```



```
/ TITLE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 11
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY      8 KXXXXXXXXXXSHPOPEK 25
Db      100 RKGWPCPSAWSHPOPEK 117

RESULT 12
US-09-977-137A-12
/ Sequence 12, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ TITLE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY      8 KXXXXXXXXXXSHPOPEK 25
Db      100 RKGWPCPSAWSHPOPEK 117

RESULT 13
US-09-977-137A-6
/ Sequence 6, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Cagliati, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ TITLE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
```

```
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 118
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match          55.9%; Score 62; DB 4; Length 118;
Best Local Similarity 44.4%; Pred. No. 0.76;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY      8 KXXXXXXXXXXSHPOPEK 25
Db      101 RKGWPCPSAWSHPOPEK 118

RESULT 14
US-09-382-906A-2
/ Sequence 2, Application US/09382906A
/ Patent No. 6448475
/ GENERAL INFORMATION:
/ APPLICANT: Dellapenna, Dean
/ APPLICANT: Shintani, David
/ TITLE OF INVENTION: Manipulation of Tocopherol Levels in Transgenic Plants
/ FILE REFERENCE: 920905.90032
/ CURRENT APPLICATION NUMBER: US/09/382,906A
/ CURRENT FILING DATE: 1999-08-25
/ PRIOR APPLICATION NUMBER: 60/097,863
/ PRIOR FILING DATE: 1998-08-25
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 318
/ TYPE: PRT
/ ORGANISM: Synechocystis PCC6803
US-09-382-906A-2

Query Match          49.5%; Score 55; DB 4; Length 318;
Best Local Similarity 38.9%; Pred. No. 31;
Matches 7; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

OY      6 FEKXXXXXXXXXXSHPOF 23
Db      209 WEKVMQOLDQSHPAF 226

RESULT 15
US-08-828-741B-11
/ Sequence 11, Application US/08828741B
/ Patent No. 6043069
/ GENERAL INFORMATION:
/ APPLICANT: Koentgen, Frank
/ APPLICANT: Sues, Gabriele M.
/ APPLICANT: Tarlinton, David M.
/ APPLICANT: Treutlein, Herbert R.
/ TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
/ TITLE OF INVENTION: PRODUCING SAME
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
/ STREET: 400 Garden City Plaza
/ CITY: Garden City
/ STATE: New York
/ COUNTRY: United States of America
/ ZIP: 11530
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/828,741B
FILING DATE: 26-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-828-741B-11

Query Match 48.6%; Score 54; DB 3; Length 155;
Best Local Similarity 34.8%; Pred. No. 16;
Matches 8; Conservative 10; Mismatches 5; Indels 0; Gaps 0;
QY 1 WSHPOFERXXXXXXXXXWSHPOF 23
DB 131 YSPGQGTKLEIKRSGSAMRHPQF 153

Search completed: March 2, 2005, 12:25:37
Job time: 28.4553 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 19.8699 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ14
Perfect score: 112
Sequence: 1 whnpqfexxxxxxxxxxwhnpqfex 26

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 183416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	64	57.1	699	2	A54660	histidine rich cal
2	59	52.7	331	2	T51894	related to nucleot
3	59	52.7	771	2	A87532	glucosylase (impo
4	56	50.0	103	2	C85713	unknown protein en
5	56	50.0	103	2	A90904	hypothetical prote
6	56	50.0	1449	2	A12017	hypothetical prote
7	55.5	49.6	2037	2	S37178	fatty-acyl-CoA syn
8	55	49.1	98	2	E84598	probable auxin-reg
9	55	49.1	985	2	D82776	pyruvate dehydroge
10	55	49.1	1030	2	T16114	hypothetical prote
11	54	48.2	103	2	E90973	hypothetical prote
12	54	48.2	103	2	A85821	unknown protein en
13	54	48.2	363	2	T26842	hypothetical prote
14	54	48.2	389	2	T44957	heat shock protein
15	54	48.2	564	2	A70383	glutamine (BC 1.
16	54	48.2	585	2	S53072	glutamate decarbox
17	54	48.2	873	2	T25442	hypothetical prote
18	54	48.2	896	2	AB1156	conserved membrane
19	54	48.2	897	2	AE1514	conserved membrane
20	54	48.2	897	2	T43628	phosphatidylinosit
21	53.5	47.8	1028	2	T16475	hypothetical prote
22	53	47.3	645	2	I50680	alpha subunit of r
23	53	47.3	667	1	A48660	transketolase (EC
24	53	47.3	667	2	G91044	transketolase 2 is
25	53	47.3	667	2	B85889	transketolase 2 is
26	52	46.4	145	2	T19489	hypothetical prote
27	52	46.4	244	2	A39365	cyanamide hydratase
28	52	46.4	351	1	OKGNSA	protein kinase (BC
29	52	46.4	352	2	B82263	conserved hypothet

30	52	46.4	452	2	F84421	hypothetical prote
31	52	46.4	731	2	T29129	probable ATP/GTP b
32	52	46.4	882	2	G83018	pyruvate dehydroge
33	52	46.4	1218	2	S71376	glutamate receptor
34	51.5	46.0	142	2	B82223	hypothetical prote
35	51.5	46.0	358	2	T32238	hypothetical prote
36	51.5	46.0	443	2	AB2719	conserved hypothet
37	51.5	46.0	480	2	P97500	tolerance to colic
38	51.5	46.0	2157	2	A13009	peptide synthetase
39	51.5	46.0	2566	2	E98274	hypothetical prote
40	51	45.5	105	2	T10227	hypothetical prote
41	51	45.5	292	2	AB1724	transcription regu
42	51	45.5	292	2	A11353	transcription regu
43	51	45.5	301	1	S42101	protein kinase (BC
44	51	45.5	315	2	S19737	hypothetical prote
45	51	45.5	320	2	AD0974	probable exported

ALIGNMENTS

RESULT 1

A54660 histidine rich calcium binding protein - human
C:Species: Homo sapiens (man)
C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 03-Jul-2004
R: Hofmann, S.L.; Topham, M.; Hsieh, C.L.; Francke, U.
Genomics 9, 656-669, 1991
A:Title: cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum protein, and its
A:Reference number: A54660; MUID:91244309; PMID:2037293
A:Accession: A54660
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-699 <HOF>
A:Cross-references: UNIPROT:P23327; GB:M60052; NID:g183918; PIDN:AAA88071.1; PID:g183915
C:Keywords: calcium binding

Query Match 57.1%; Score 64; DB 2; Length 699;
Best Local Similarity 37.5%; Pred. No. 4.4;
Matches 9; Conservative 10; Mismatches 5; Indels 0;
Gaps 0;
QY 3 HPQFKXXXXXXXXXXSHPOFEK 26
DB 67 HPDENKDVSTENGHHFMSHPDREK 90

RESULT 2

T51894 related to nucleotide-binding protein [imported] - Neurospora crassa
N:Alternate names: protein B23111.60
C:Species: Neurospora crassa
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
R:Schulze, U.; Aigner, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
C:Accession: U; T51894
A:Title: The Protein Sequence Database, August 2000
A:Reference number: Z25858
A:Accession: T51894
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <SCH>
A:Cross-references: EMBL:AL391572; GSPDB:GN00116; NCSP:B23111.60
A:Experimental source: BAC clone B23111; strain OR74A
C:Genetics:
A:Gene: NCSP:B23111.60
A:Map position: 6
A:Introns: 34/3; 60/3; 87/1
C:Superfamily: conserved probable membrane protein Y1L003W

```
Query Match Similarity      52.7%   Score 59; DB 2; Length 331;
Best Local Similarity      28.0%   Pred: No. 9.1;
Matches       7; Conservative    11; Mismatches       7; Indels     0; Gaps     0;
```

RESULT 3
A87532
glucosylase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001
C:Accession: A87532
#text_change 09-Jul-2004

R.; Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Debov, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon-
 nov, N.J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.;
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001.
 A>Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A|Reference number: A874249; MUID:21173698; PMID:11259647
 A|Accession: A87532
 A|Status: preliminary
 A|Molecule type: DNA
 A|Residues: 1-771 <STO>
 A|Cross-references: NIMPROT:Q9A613; GB:AE005673; NID:gl3423797; PIDN:AAK424253.1; GSPDB:G-
 C|Genetics:
 C|Gene: CC282
 C|Superfamily: Clostridium glucan 1,4-alpha-glucosidase

RESULT 4
C85713
unknown protein encoded within prophage CP-9330 [imported] - *Escherichia coli* (strain O157:H7)
C/Species: *Escherichia coli*
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/accession: C85713
R/Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Agodcas
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A/Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: C85113
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-103 <SIO>
A:Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:g12515046; PIDN:AA656167.1; GSPDB:Q
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2097

```
Query Match      50.0%; Score 56; DB 2; Length 103;
Best Local Similarity 22.7%; Pred. No. 6.1;
Matches          5; Conservative 16; Mismatches 1; Indels 0; Gaps 0;
```

```

RESULT 5
A090904
hypothetical protein Eca2201 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

```

C/Accession: A90904
R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G.
G.Sawaya, N.; Yesunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genetic
A/Reference number: A99629; NCID:21156231; PMID:11258796
A/Accession: A90904

A: molecule type: DNA
A: Residues: 1-103 <HAY>
A: Cross-references: UNIPROT:Q8X549; GB:BA000007; PDB:1AB35624.1; PID:913361667; GSPDB:G
A: Experimental source: strain O157:H7, substrain R1MD 0509952
C: Genetics:
A: Gene: EC92201

Query	Match	50.0%	Score	56	DB	2	Length	103
	Best	51.1%						
	Local	51.1%	Pred.	No.	6.1			
	Matches	5	Conservative	16	Mismatches	1	Indels	0
								Gaps 0
Qy	5	QPEKXXXXXXXXXXWSHPQEK	26					
							
							
Db	3	ETKEKMPVVRDGGYWTHTPEYK	24					

RESULT 6
AT2017
hypochemical protein all1695 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: *Nostoc* sp. PCC 7120
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AT2017
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriduchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium *Anabaena* strain PCC 7120
R:Reference number: AB1807; MUID:21595285; PMID:11759840
C:Accession: AT2017

```

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 17449 <KUR>
A;Cross-references: UNIPROT:08YWC0, GB:BA000019, FID:BAB78061, PID:G17135515, GSPDB:C
A;Experimental source: strain FCC 7120
C;Genetics:
X;Gene: all1695

```

Query Match	50.0%;	Score 56;	DB 2;	Length 1449;
Best Local Similarity	28.6%;	Pred. No. 1.3e+02;		
Matches	6;	Conservative 12;	Mismatches 3;	Indels 0;
				Gaps 0;
Qy	1	WSHPQFEKXXXXXXXXXXWSH	21	
Db	809	WHQQQFOQPEVYTLPLVEYWSH	829	

RESULT 7
S37178
fatty-acyl-CoA synthase (EC 2.3.1.86) beta chain - yeast (Candida albicans)
N/Alternate names: fatty-acid synthase beta chain
C/Species: Candida albicans
C/Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C/Accession: S37178
R/Zhao, X.J.; Cihlar, R.L.
submitted to the EMBL Data Library, September 1993
A/Description: Isolation and sequence analysis of the Candida albicans FAS1 gene.

A/Reference number: S37178
A/Accession: S37178
A/Molecule type: DNA
A/Residues: 1-2037 <ZHA>
A/Cross-references: UNIPROT:P34731, EMBL:X74952, NID:g402176, PIDD:CAA52907.1, PID:g4021
C/Genetics
A/Gene: FAS1
C/Superfamily: yeast fatty-acyl-CoA synthase beta chain
C/Keywords: acyltransferase; coenzyme A

```
Query Match          49.6%; Score 55.5; DB 2; Length 2037;
Best Local Similarity 31.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 12; Mismatches 2; Indels 1; Gaps 1;

QY 3 HPOFEKXXXXXXXXXX-SHPQ 23
|||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:
Db 189 HPSFDKITYGILNLSMLKHP 210

RESULT 8
E84539
probable auxin-regulated protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: E84539
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.;
euser, D.; Mierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84539
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-98 <STO>
A:Cross-references: UNIPROT:Q9SKP2; GB:AE002093; NID:g4603922; PIDN:AAD29795.1; GSPDB:GN
C:Genetics:
A:Gene: At2g21210
A:Map position: 2
C:Superfamily: auxin-induced protein 10A

Query Match          49.1%; Score 55; DB 2; Length 98;
Best Local Similarity 33.3%; Pred. No. 8;
Matches 7; Conservative 12; Mismatches 2; Indels 0; Gaps 0;

QY 2 SHPOFEKXXXXXXXXXXMSHP 22
|||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:
Db 54 SHPOFKILRKAEERGFDPH 74

RESULT 9
D82776
pyruvate dehydrogenase XPO669 [imported] - Xylella fastidiosa (strain 9asc)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: D82776
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82776
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-985 <SIM>
A:Cross-references: UNIPROT:Q9PF18; GB:AE003911; GB:AE003849; NID:g9105548; PIDN:AAF8347
A:Experimental source: strain 9asc
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Porty, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Author: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
U.D.; Junqueira, M.L.; Kemper, E.L.; Klajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigi
chado, M.A.; Madeira, A.M.B.N.; Madureira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Author: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.R.; de Sa, R.G.; Sancelli, R.V.; Sawasak
A:Author: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.W.; Silva Jr., W.A.; da Silveir
M.; Ibanako, M.H.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
```

```
A:Gene: XF0669
C:Superfamily: pyruvate dehydrogenase (lipoamide); thiamin pyrophosphate-binding domain

Query Match          49.1%; Score 55; DB 2; Length 985;
Best Local Similarity 29.6%; Pred. No. 1.2e+02;
Matches 8; Conservative 12; Mismatches 5; Indels 2; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXS--HPOFE 25
|||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:
Db 854 WSCPSFSELRKDGFDTERMNLHPDSE 880

RESULT 10
T16114
hypothetical protein F20D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16114
R:Wu, X.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F20D12.
A:Reference number: Z18462
A:Accession: T16114
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1030 <MTX>
A:Cross-references: UNIPROT:Q19645; EMBL:U40933; NID:g1072142; PID:g1072148; PIDN:AAA816
C:Genetics:
A:Gene: CESP:F20D12.1
A:Introns: 159/1; 212/2; 263/3; 334/2; 385/1; 513/3; 754/2; 787/3; 906/3; 969/3; 999/2

Query Match          49.1%; Score 55; DB 2; Length 1030;
Best Local Similarity 33.3%; Pred. No. 1.2e+02;
Matches 7; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

QY 4 POFEKXXXXXXXXXXMSHPQF 24
|||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:
Db 299 PDLKPSFVQITLNMSPDRF 319

RESULT 11
E90973
hypothetical protein E90973 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90973
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gesawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAY>
A:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BNB36180.1; PID:g13362225; GSPDB:C
A:Experimental source: substrain R1MD 0509952
C:Genetics:
A:Gene: E90973

Query Match          48.2%; Score 54; DB 2; Length 103;
Best Local Similarity 25.0%; Pred. No. 12;
Matches 5; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

QY 7 EKXXXXXXXXXXSHPOFEK 26
|||||:|||||:|||||:|||||:
|||||:|||||:|||||:|||||:
Db 5 KEMPVRRNRYGCMTHPEYRK 24

RESULT 12
A85821
unknown protein encoded within prophage CP-931U [imported] - Escherichia coli (strain O1
C:Species: Escherichia coli
```

C|Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C|Accession: AB58821
R|Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Diallanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-535, 2001
A|Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A|Reference number: AB5480; MUID:21074935; PMID:11206551
A|Accession: AB58821
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-103 <STQ>
A|Cross-references: UNIPROT:O8X4V0; GB:AE005174; NID:G12516136; PIDN:AA657029.1; GSPDB:G
A|Experimental source: strain O157:H7, substrain EDL933
A|Genetics:
A|Gene: Z3120

Query Match	48.2%	Score 54	DB 2	length 103
Best Local Similarity	25.0%	Pred. No. 12		
Matches	5	Conservative	15	Mismatches
			0	Indels
				Gaps
Qy	7	XXXXXXXXXXXXXSHPOREK	26	
	 : : :		
Db	5	KEMPEVNERNEYGWTIPREYK	24	

RESULT 13
T26842
hypothetical protein Y43F4B.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Aug-2004
C:Accession: T26842
R:Matthews, L.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z20276
A:Accession: T26842
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-363 <WIL>
A:Cross-references: UNIPROT:Q45933; EMBL:AL021481; PIDD:CAA16333.1; CESP:Y43F4B.4
A:Experimental source: clone Y43F4B
C:Genetic:
A:Gene: CESP:Y43F4B.4
A:Introns: 22/3, 59/3, 80/2, 103/1, 232/2, 318/1
C:Superfamily: WD repeat homology

```

Query Match      48.2%   Score 54; DB 2; Length 363;
Best Local Similarity 25.0%   Pred. No. 51,
Matches 6; Conservative 12; Mismatches 6; Indels 0; Gaps 0.

QY      1 N$HPOPEKXXXXXXXXXXSHPOF 24
      | : : : : : : : : : : : :
Db      52 MRRSAHKCHGAVRVIMAHPEF 75

```

```

RESULT 14
T44957
heat shock protein dnaJ [similarity] - Halobacterium salinarum (ATCC 33170)
N:Alternate names: 40K chaperone, 40K heat shock protein
C:Species: Halobacterium salinarum
A:Variety: ATCC 33170
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T44957
R:Busstard, K.; Gupta, R.S.
J. Mol. Evol. 45, 193-205, 1997
A:Title: The sequences of heat shock protein 40 (DnaJ) homologs provide evidence for a c
A:Reference number: Z22880; MUID:97383250; PMID:9236279
A:Accession: T44957
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-389 <BUS>
A:Cross-references: UNIPROT:O34135; EMBL:U093357; PIDDN:AA896891.1
A:Experimental source: ATCC 33170
A>Note: the source is designated as Halobacterium cutribdrium

```

C:Genetics:
A:Gene: dnaJ
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
P:4-68/Domain: dnaJ amino-terminal homology <DNU>

Query Match	48.2%	Score 54	DB 2	Length 389
Best Local Similarity	33.3%	Pred. No. 55		
Matches	7	Conservative	12	Mismatches 2; Indels 0; Gaps 0
QY	3	HPDFEKKXXXXXXXXXMSHPQ	23	
DB	276	HPDFEEDDTTSITATPFSFPQ	286	

RESULT 15
A70383
hydrogenase (EC 1.18.99.1) 2 large chain - Aquifex aeolicus
CISpecies: Aquifex aeolicus
CDate: 08-May-1998 #sequenc_revision 08-May-1998 #text_change 09-Jul-2004
CAccession: A70383
RDecker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:9819666; PMID:9537320
Accession: A70383

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-564 <ADP>
A:Cross-references: UNIPROT:O67092; GB:AE000715; NID:g2983460; PIDN:AAC07046.1; PID:g298
A:Experimental source: strain VF5
C:Genetics:
A:Gene: mbhL2
C:Function:
A:Pathway: hydrogen metabolism
A:Note: contains iron-sulfur and nickel
C:Superfamily: hydrogenase (NiFe) large chain
C:Keywords: hydrogen metabolism; iron; iron-sulfur protein; membrane bound; metalloprote
E:60.63,546,549/binding site: nickel (Cys) #status predicted
E:63,549/binding site: iron (Cys) #status predicted
E:67/Active site: His #status predicted

Query Match	48.2%;	Score 54;	DB 2;	Length 566;
Best Local	Similarity	23.8%;	Pred. No. 85;	
Matches	5;	Conservative	15;	Mismatches
			1;	Indels
			0;	Gaps
QY	5	QFEKXXXXXXXXXXMSHPDE	25	
Db	157	QVGRGQGLGIPNKGVMNHPYK	177	

Search completed: March 2, 2005, 12:29:00
Job time : 20.8699 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 96.6016 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ14
Perfect score: 112
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfek 26

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	57.1	699	1 SRCH HUMAN	P23327 homo sapien
2	62	55.4	722	2 Q7YSZ1	Q7YSZ1 maccia fasc
3	61	54.5	704	2 Q8CJD2	Q8CJD2 rattus norv
4	61	54.5	743	2 Q8CH90	Q8CH90 rattus norv
5	60	53.6	782	2 Q9HGH1	Q9HGH1 picchia guil
6	59	52.7	253	2 Q72NB4	Q72NB4 leprospira
7	59	52.7	253	2 Q8F8C3	Q8F8C3 leprospira
8	59	52.7	309	2 Q7S6P7	Q7S6P7 neurospora
9	59	52.7	400	2 Q8TTC4	Q8TTC4 methanosarc
10	59	52.7	771	2 Q9A613	Q9A613 xanthomonas
11	58	51.8	895	2 Q8P4T1	Q8P4T1 xanthomonas
12	57.5	51.3	107	2 Q6Z315	Q6Z315 oryza sativ
13	57.5	51.3	380	2 Q742U3	Q742U3 mycobacteri
14	57	50.9	202	2 Q6E071	Q6E071 oryza sativ
15	57	50.9	1111	2 Q7SAB0	Q7SAB0 neurospora
16	56.5	50.4	270	2 Q9BLC4	Q9BLC4 leishmania
17	56	50.0	103	2 Q8X549	Q8X549 escherichia
18	56	50.0	192	2 Q7VBE9	Q7VBE9 prochloroco
19	56	50.0	299	2 Q9A6T7	Q9A6T7 rhizobium 1
20	56	50.0	1449	2 Q7XNS9	Q7XNS9 oryza sativ
21	56	50.0	508	1 Q7XNS9	Q7XNS9 oryza sativ
22	56	50.0	508	1 Q9B8G0	Q9B8G0 marichium s
23	56	50.0	508	2 Q9B8G3	Q9B8G3 oserya coul
24	56	50.0	577	2 Q8EXI8	Q8EXI8 homo sapien
25	56	50.0	640	1 HTPG_NITRU	HTPG_NITRU
26	56	50.0	1449	2 Q8YWC0	Q8YWC0 nitrososoma
27	55.5	49.6	409	2 Q7QHY8	Q7QHY8 anopheles g
28	55.5	49.6	2037	1 FAST_CANAL	FAST_CANAL
29	55	49.1	98	2 Q9SKP2	Q9SKP2 arabidopsis
30	55	49.1	181	2 Q8MXH8	Q8MXH8 plodia inte
31	55	49.1	295	2 Q949H3	Q949H3 hevea bras

32	55	49.1	295	2 Q8GUD7	Q8GUD7 hevea bras
33	55	49.1	354	2 Q8Z2N5	Q8Z2N5 pyrobaculum
34	55	49.1	381	2 Q7Q9L2	Q7Q9L2 anopheles g
35	55	49.1	847	2 Q8EEC3	Q8EEC3 shewanella
36	55	49.1	985	2 Q9PFI8	Q9PFI8 xylella fas
37	55	49.1	1030	2 Q19645	Q19645 caenorhabdi
38	55	49.1	4903	1 MLJ3_MOUSE	MLJ3_MOUSE
39	55	49.1	4911	1 MLJ3_MOUSE	MLJ3_MOUSE
40	54.5	48.7	334	2 Q6ZB50	Q6ZB50 homo sapien
41	54.5	48.7	334	2 Q6ZB50	Q6ZB50 burkholderi
42	54.5	48.7	471	2 Q8NTJ3	Q8NTJ3 corynebacte
43	54.5	48.7	663	2 Q8J219	Q8J219 aspergillus
44	54	48.2	103	2 Q8X4V0	Q8X4V0 escherichia
45	54	48.2	208	2 Q6ZTR3	Q6ZTR3 bacillus 11

ALIGNMENTS

RESULT 1
ID SRCH_HUMAN STANDARD, PRT, 699 AA.
AC P23327;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Sarcoplasmic reticulum histidine-rich calcium-binding protein
DE Precursor.
GN Name=HRC; Synonyms=HCP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo;
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC TISSUE=Skeletal muscle;
RX MEDLINE=9124309; PubMed=2037293;
RA Hofmann S.L., Topham M., Hsieh C.-L., Franke U.;
RT "cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum
RT protein, and localization of the gene to human chromosome 19 and mouse
RT chromosome 7."
RL Genomics 9:656-669(1991).
CC -!- FUNCTION: May play a role in the regulation of calcium
CC sequestration or release in the SR of skeletal and cardiac muscle.
CC -!- SUBCELLULAR LOCATION: Sarcoplasmic reticulum lumen.
CC -!- SIMILARITY: Strong, to rabbit HRC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, ME0052; AAA8071.1; -;
DR PIR: A54660; A54660.
DR GeneW: HGNC:5178; HRC.
DR MIM: 142705; -;
DR GO: GO:0005509; P:calcium ion binding; TAS.
DR GO: GO:0006936; P:muscle contraction; TAS.
KW Calcium-binding; Polymorphism; Repeat; Signal.
FT SIGNAL 1 28
FT CHAIN 29 699
FT
FT DOMAIN 193 204 Sarcoplasmic reticulum histidine-rich
FT DOMAIN 246 261 calcium-binding protein.
FT DOMAIN 106 365 Glu-rich (acidic).
FT REPEAT 180 213 4 x tandem repeats, acidic.
FT REPEAT 238 270 1-1.
FT REPEAT 295 318 1-2.
FT REPEAT 343 365 1-3.
FT DOMAIN 106 342 6 x approximate tandem repeats.

```

FT REPEAT 106 121 2-1.
FT REPEAT 134 154 2-2.
FT REPEAT 155 177 2-3.
FT REPEAT 214 237 2-4.
FT REPEAT 271 294 2-5.
FT REPEAT 319 342 2-6.
FT DOMAIN 627 673 Metal-binding (potential).
FT VARIANT 96 96 S -> A (in dbSNP:3745287).
FT VARIANT 204 204 /FTID=VAR_005623.
FT VARIANT 204 204 /FTID=VAR_011622.
SQ SEQUENCE 699 AA; 80244 MW; 9922EEDF012C61D CRC64;

Query Match
Best Local Similarity 57.1%; Score 64; DB 1; Length 699;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 3 HPOFEKXXXXXXSHPOFEK 26
DB 67 HPDENKDVSTENGHFWSHPDREK 90

RESULT 2
Q7YS21 PRELIMINARY; PRT; 722 AA.
AC Q7YS21;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Histidine-rich calcium binding protein.
GN Name=HRC;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Hong S., Cho C.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; A305856; AAF70486.1; -.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; UNKNOWN_1.
DR PROSITE; PS00328; HCP; UNKNOWN_1.
SQ SEQUENCE 722 AA; 82753 MW; 87CBE737FA647F24 CRC64;

Query Match
Best Local Similarity 55.4%; Score 62; DB 2; Length 722;
Matches 8; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

QY 2 SHPOFEKXXXXXXSHPOFEK 26
DB 66 NHDPGNKDVSTENGHFWSHPDREK 90

RESULT 3
Q8CJD2 PRELIMINARY; PRT; 704 AA.
AC Q8CJD2;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Guanylyl cyclase alpha 1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura I., Suzuki N.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AB096020; BAC24016.1; -.
DR HSSP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR01054; G_cyclase.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match
Best Local Similarity 54.5%; Score 61; DB 2; Length 704;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXSHPOFEK 26
DB 687 SGVDLVRGSMWSHPDREK 704

RESULT 4
Q8CH90 PRELIMINARY; PRT; 743 AA.
AC Q8CH90;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Nakamura I., Yao Y., Suzuki N.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB097860; BAC44887.1; -.
DR HSSP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR001054; G_cyclase.
DR InterPro; IPR009080; ERNabyn_1a_bind.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match
Best Local Similarity 54.5%; Score 61; DB 2; Length 743;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXSHPOFEK 26
DB 726 RETSLVPRGSMWSHPDREK 743

RESULT 5
Q9GH1 PRELIMINARY; PRT; 782 AA.
AC Q9GH1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Type II DNA topoisomerase (Fragment).
GN Name=top2;
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxId=4929;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21363449; PubMed=11470534; DOI=10.1016/S0378-1119(01)00526-1;
RA Kato M., Oeki M., Kikuchi A., Kanbe T.;
RT "Phylogenetic relationship and mode of evolution of yeast DNA
topoisomerase II gene in the pathogenic Candida species.";
RL Gene 272:275-281(2001).
CC -1- FUNCTION: Control of topological states of DNA by transient
breakage and subsequent rejoining of DNA strands. Topoisomerase II
makes double-strand breaks (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
of double-stranded DNA.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- MISCELLANEOUS: Eukaryotic topoisomerase I and II can relax both
negative and positive supercoils, whereas prokaryotic enzymes
relax only negative supercoils (By similarity).
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
DR EMBL: AB049145; BAB13754.1; -.
DR HSP: P06786; IBDT.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0005524; P:ATP binding; IEA.
DR GO: GO:0003918; P:DNA topoisomerase (ATP-hydrolyzing) activity; IEA.
DR GO: GO:0006265; P:DNA topological change; IEA.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR003957; CBFA_NFYB_topis.
DR InterPro: IPR011558; DNA_gyrase_B.
DR InterPro: IPR012411; DNA_topoisomI.
DR InterPro: IPR022205; DNA_topoisomIV.
DR Pfam: PF00204; DNA_gyraseB_1.
DR Pfam: PF00521; DNA_topoisomIV_1.
DR Pfam: PF02518; HATPase_C_1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PRINTS: PR00418; TP12FAMILY.
DR ProDom: PD149633; DNA_gyrase_B_1.
DR ProDom: PD000742; DNA_topoisomIV_1.
DR SMART: SM00387; HATPase_C_1.
DR SMART: SM00434; TOP2c_1.
DR SMART: SM00387; HATPase_C_1.
DR ATP-binding; DNA-binding; Isomerase; Topoisomerase.
KW NON_TER
FT NON_TER
SQ SEQUENCE 782 AA; 88612 MW; 0B9CFDB3BDAA6D CRC64;

Query Match
Best Local Similarity 28.0%; Score 60; DB 2; Length 782;
Matches 7; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXWSHPDF 25
DB 544 YNMPFEKWRTEGTCTCKWKHYFK 568

RESULT 6
Q72N84 PRELIMINARY; PRT; 253 AA.
ID Q72N84
AC Q72N84
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Dipeptide/oligopeptide ABC transport system permease protein.
GN OrderedlocusNames=LOC12953;
OS Leposiphira interorgans (serogroup Icterohaemorrhagiae / serovar
Copenhagense).
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxId=44275;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fluctuz L1-130;
PubMed=15028702; DOI=10.1128/JB.186.7.2164-2172.2004;
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Montelero-Vicorello C.B.,

```

```

RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartekerk R.A.,
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H.,
RA Coutinho L.L., Degraeve W.M., Dellagostin O.A., El-Dorry H.,
RA Ferro E.S., Ferro M.I.T., Furian L.R., Gamberini M., Gigliotti E.A.,
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Hasekava R.T.,
RA Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,
RA Kuramae E.E., Lemos B.G.M., Lemos M.V.F., Martino C.L., Nunes L.R.,
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schiefer A.,
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
RA Camargo L.B.A., Kitajima J.P., Secubal J.C., Van Sluys M.A.;
RT "Comparative genomics of two Leptospira interrogans serovars reveals
RT novel insights into physiology and pathogenesis.";
RL J. Bacteriol. 186:2164-2172(2004).
CC -1- FUNCTION: Part of a binding-protein-dependent transport system.
Probably responsible for the translocation of the substrate across
the membrane (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the binding-protein-dependent transport
system permease family.
DR EMBL: AE017298; AAS71503.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp_1; 1.
DR PROSITE: PS50928; ABC_TM1_1.
KW Complete proteome; Transmembrane; Transport.
SQ SEQUENCE 253 AA; 27677 MW; 0A145ACBE01964C CRC64;

Query Match
Best Local Similarity 31.8%; Score 59; DB 2; Length 253;
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXWSHP 22
DB 22 WKNPTEVLEDSFCSVTWSP 43

RESULT 7
Q8F8C3 PRELIMINARY; PRT; 253 AA.
ID Q8F8C3
AC Q8F8C3
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Dipeptide/oligopeptide/nickel transport systems, permease
components.
GN OrderedlocusNames=LA0634;
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxId=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
MEDLINE=22598143; PubMed=12712204; DOI=10.1038/nature01597;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia U., Tu Y.-F.,
RA Zhang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-Y., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "unique physiological and pathogenic features of Leptospira
interrogans revealed by whole-genome sequencing.";
RL Nature 422:888-893(2003).
CC -1- FUNCTION: Part of a binding-protein-dependent transport system.
Probably responsible for the translocation of the substrate across
the membrane (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the binding-protein-dependent transport
system permease family.
DR EMBL: AE011250; AAN47833.1; -.
DR GO: GO:0016021; C:integral to membrane; IEA.

```

```

DR GO:0005215; F:transporter activity; IEA.
DR GO:0006810; P:transport; IEA.
DR InterPro:IPR000515; BPD_transp.
DR Pfam:PF00528; BPD_transp_1; 1.
DR PROSITE:PS50928; ABC_TM1; 1.
DR Complete proteome; Transmembrane; Transport.
SQ SEQUENCE 253 AA; 27677 MW; 0A145ACBE801964C CRC64;

Query Match
Best Local Similarity 52.7%; Score 59; DB 2; Length 253;
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHP 22
Db 22 WKNPTEVFLEDFSCVTWSHP 43

RESULT 8
ID 0756P7 PRELIMINARY; PRT; 309 AA.
AC 0756P7;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DE Hypothetical protein.
GN Name=NCU04788.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetiales; Sordariales; Sordariaceae; Neurospora.
OC NCBI_TaxId=5141;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Putcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Seltzer G.O., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kolbe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gierre S.,
RA Kamel M., Kamyselis M., Mauceli E., Bielke C., Rudd S., Fishman D.,
RA Kytrofova S., Raamsen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macho G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA Desouza C.C., Glaes L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Naveg O., Alex L.A., Manhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RA "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RT Nature 0:0-0(2003).
RL -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL: AABX01000323; EAA31178.1; -
DR EMBL: AABX01000323;
KW Hypothetical protein.
SQ SEQUENCE 309 AA; 32912 MW; ABE6A0CD76802891 CRC64;

Query Match
Best Local Similarity 52.7%; Score 59; DB 2; Length 309;
Matches 7; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHP 25
Db 137 WKGPMILKAIQQLLHEVDMSHPSLD 161

RESULT 9
ID 08TK4 PRELIMINARY; PRT; 400 AA.
AC 08TK4;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Aldo/keto reductase.

```

```

GN OrderedLocusNames=MA0427;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxId=2214;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932368; DOI=10.1101/gr.223902;
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Strange-Thomann N., Deatellano K., Johnson R.,
RA Linton L., McSwan P., McKernan K., Talmas J., Titrrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Gramam D.E., Gramme D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AEO10703; AAM03874.1; -
DR HSSP: P52895; 1396.
DR GO: GO:0005449; F:electron transporter activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR Pfam: PF00248; Aldo_ket_red; 1.
DR Pfam: PF00037; Fer4; 1.
DR PRINTS: PR00069; ALDREDTASE.
DR PRODOM: PD000288; Aldo/ket_red; 1.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 1.
KW 4fe-4s; Complete proteome; iron; iron-sulfur; Metal-binding.
SQ SEQUENCE 400 AA; 45143 MW; 7PAD1773FE7DC29 CRC64;

Query Match
Best Local Similarity 52.7%; Score 59; DB 2; Length 400;
Matches 6; Conservative 13; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSHP 23
Db 222 WSEADVERTPEVMAIRWVNHPE 244

RESULT 10
ID 09A613 PRELIMINARY; PRT; 771 AA.
AC 09A613;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Glucosylase.
GN OrderedLocusNames=CC2282;
OS Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteriaceae; Caulobacter.
OC NCBI_TaxId=155892;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Berry K.J., Utecherback T.R., Tran K., Wolf A.M., Vamathevan J.J.,
RA Ermolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
RA Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AEO05898; AAK24253.1; -

```

DR PIR: A87532; A87532.
 DR HSP: O85672; 1LP6.
 DR TIGR: CC2282; 1LP6.
 DR InterPro: IPR01013; Gal_mut_like.
 DR InterPro: IPR006425; Glucan_glucoald.
 DR InterPro: IPR008928; Glyco_trans_6hp.
 DR TIGRFAMs: TIGR01535; glucan_glucoald; 1.
 KW Complete proteome.
 SQ SEQUENCE 771 AA; 82468 MW; D554EC477C4B4409 CRC64;

Query Match 52.7%; Score 59; DB 2; Length 771;
 Best Local Similarity 28.6%; Pred. No. 1.9e+02;
 Matches 6; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXWSH 21
 Db 741 WSHAETVXKLRESDGQVWDH 761

RESULT 11

Q8P4T1 PRELIMINARY; PRT; 895 AA.
 AC Q8P4T1;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Pyruvate dehydrogenase.
 GN Name=aceE; OrderedLocNames=XCC3625;
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 CX NCBI_Taxid=340;
 RX SEQUENCE FROM N.A. / NCPPB 528;
 RC STRAIN=ATCC 33913; PubMed=12024217; DOI=10.1038/417459a;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Ounggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Camnava F., Cardoso J., Chambergo J.R., Clapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Fortiglieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locelli E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kiteajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463 (2002).
 DR EMBL: AE012483; AAM42895.1; -
 DR HSRP: P06958; 1LBA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR InterPro: IPR004660; ACEB.
 DR InterPro: IPR005474; Transketolase N.
 DR InterPro: IPR009014; Transketo C like.
 DR Pfam: PF00456; Transketolase_N_I.
 DR TIGRFAMs: TIGR00759; aceE; 1.
 KW Complete proteome.
 SQ SEQUENCE 895 AA; 99758 MW; F55EC175A86C73AF CRC64;

Query Match 51.8%; Score 59; DB 2; Length 895;
 Best Local Similarity 28.6%; Pred. No. 3e+02;
 Matches 8; Conservative 14; Mismatches 4; Indels 2; Gaps 1;

Qy 1 WSHPOFEKXXXXXXXXXWSH--HPOFEK 26
 Db 764 WSCPSFNEILRDGFDALERNRLHPEAEQ 791

RESULT 12

ID Q6Z315 PRELIMINARY; PRT; 107 AA.
 AC Q6Z315;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE DCL protein-like.
 GN Name=OJ1004_A11.16-2; Synonym=P0539D10.35-2;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrharioideae; Oryzoae; Oryza.
 CX NCBI_Taxid=39947;
 RX SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP005287; BAD17315.1; -
 DR EMBL: AP004817; BAD17127.1; -
 SQ SEQUENCE 107 AA; 12839 MW; 00F06AF3A12B0EB1 CRC64;

Query Match 51.3%; Score 57.5; DB 2; Length 107;
 Best Local Similarity 37.5%; Pred. No. 35;
 Matches 9; Conservative 12; Mismatches 2; Indels 1; Gaps 1;

Qy 3 HPOFEKXXXXXXXXXWSH--HPOFE 25
 Db 30 HPOYKKGICGIDYLVGLHPEFE 53

RESULT 13

ID Q742U3 PRELIMINARY; PRT; 380 AA.
 AC Q742U3;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein
 GN OrderedLocNames=MAP0742c;
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_Taxid=1770;
 RX SEQUENCE FROM N.A.
 RP STRAIN=K10;
 RA Li L., Bannantine J., Zhang Q., Amosin A., Alt D., Kapur V.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE017229; AAS03059.1; -
 DR GO: GO:0001760; F:aminocarbonyl-semialdehyde decarboxy. .; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 KW Complete proteome.
 SQ SEQUENCE 380 AA; 44168 MW; 2BA1AB05F50B71FE CRC64;

Query Match 51.3%; Score 57.5; DB 2; Length 380;
 Best Local Similarity 23.1%; Pred. No. 1.4e+02;
 Matches 6; Conservative 15; Mismatches 4; Indels 1; Gaps 1;

Qy 1 WS-HPOFEKXXXXXXXXXWSHPOFE 25
 Db 162 WAGHYPMQILIKAPRPSWGHPRXD 187

RESULT 14

ID Q6E071 PRELIMINARY; PRT; 202 AA.
 AC Q6E071;
 Q6E071;

DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein OSJNB0017118.11 (Hypothetical protein
 DE OSJNB0095104.17).
 CN Name=OSJNB0017118.11; Synonym=OSJNB0095104.17;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 OC NCBI_TaxId=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Katayose Y.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005724; BAD29199.1; -.
 DR EMBL; AP005701; BAD29131.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 202 AA; 22026 MW; 23392FDF138284FB CRC64;

Query Match 50.9%; Score 57; DB 2; Length 202;
 Best Local Similarity 27.3%; Pred. No. 81;
 Matches 6; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHP 22
 DB 148 WGHPSFSGRCRCRRRRAHP 169

RESULT 15
 ID Q7SAB0 PRELIMINARY; PRT; 1111 AA.
 AC Q7SAB0;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Predicted protein.
 GN Name=NCU06321.1;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OC NCBI_TaxId=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OR74A;
 RA Galagan J.E., Borkovich K.A., Selker E.U., Read N.D.,
 RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehm B.,
 RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
 RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
 RA Selkrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
 RA Kotne G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
 RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
 RA Kamal M., Kamysaselis M., Mauceli E., Bielke C., Rudd S., Fishman D.,
 RA Kryzofova S., Rasmussen C., Metzger R.L., Perkins D.D., Kroken S.,
 RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Omani S.A.,
 RA Desouza C.C., Glas L., Orbach M.J., Berglund J., Voelker R.,
 RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
 RA Natvig D.O., Alex L.A., Manhaupt G., Ebbole D.O., Freitag M.,
 RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
 RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
 RL Nature 0:0-0 (2003).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC DR EMBL; AABX0100205; EAA33347.1; -.
 SQ SEQUENCE 1111 AA; 121374 MW; 31264510AF33B363 CRC64;

Query Match 50.9%; Score 57; DB 2; Length 1111;
 Best Local Similarity 28.0%; Pred. No. 5.1e+02;
 Matches 7; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHP 25
 DB 148 WGHPSFSGRCRCRRRRAHP 169

DB 292 WGHDDVVEWQHDPDVEWQHDPDE 316

Search completed: March 2, 2005, 12:44:29
 Job time : 97.6016 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 117.106 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ14
Perfect score: 112
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfek 26

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16dec04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20048:*
8: geneseqp20058:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	99	88.4	24	6	ABP60362	Abp60362 Streptavir
2	97.5	87.1	35	6	ABP60369	Abp60369 Streptavir
3	97	86.6	36	6	ABP60370	Abp60370 Streptavir
4	72	64.3	236	6	AD019051	Ad019051 Murine an
5	72	64.3	236	8	AD019053	Ad019053 Murine an
6	70	62.5	24	6	ABP60363	Abp60363 Streptavir
7	68	60.7	209	6	ADA00709	Ada00709 Modified
8	68	60.7	258	6	ADA27293	Ada27293 Plasmid f
9	68	60.7	258	6	ADA00702	Ada00702 Modified
10	68	60.7	400	6	ADA27291	Ada27291 Plasmid f
11	68	60.7	400	6	ADA27292	Ada27292 Plasmid f
12	68	60.7	400	6	ADA00700	Ada00700 Modified
13	68	60.7	400	6	ADA00707	Ada00707 Modified
14	68	60.7	400	6	ADA00701	Ada00701 Modified
15	66	58.9	646	8	ADS20251	Ads20251 Human ags
16	66	58.9	858	8	ADS20230	Ads20230 Human ags
17	65	58.0	539	6	ABG74883	Abg74883 Bacterioph
18	65	58.0	539	8	AD139158	Ad139158 Enterobac
19	64.5	57.6	485	8	ADS20249	Ads20249 Furin-pri
20	64.5	57.6	697	8	ADS20227	Ads20227 Human ags
21	64	57.1	699	7	ADJ68628	Adj68628 Human hea
22	64	57.1	699	8	ADQ18833	Adq18833 Human sol
23	64	57.1	699	8	ADQ39188	Adq39188 Human myo
24	64	57.1	699	8	ADQ39189	Adq39189 Human myo
25	63	56.2	183	6	ADA27294	Ada27294 Plasmid f

26	63	56.2	183	6	ADPA00703	Adp00703 Modified
27	63	56.2	467	8	ADP18461	Adp18461 Arthropod
28	62	55.4	56	4	AAE04104	AAe04104 Human ge
29	62	55.4	56	4	AAE04132	AAe04132 Human ge
30	62	55.4	56	5	ABG64354	ABg64354 Human al
31	62	55.4	56	5	ABG64356	ABg64356 Human al
32	62	55.4	56	6	ADA98358	Ada98358 Human se
33	62	55.4	56	6	ADA44167	Ada44167 Human se
34	62	55.4	56	7	ADC20522	Adc20522 Human se
35	62	55.4	56	7	ADP10783	Adp10783 Human se
36	62	55.4	56	8	ADL77621	Adl77621 Albumin
37	62	55.4	56	8	ADL77623	Adl77623 Albumin
38	62	55.4	57	4	AAU01630	AAu01630 Human se
39	62	55.4	267	8	ADR70320	Adr70320 Poliovir
40	62	55.4	268	8	ADR70385	Adr70385 Poliovir
41	62	55.4	268	8	ADR70319	Adr70319 Poliovir
42	62	55.4	268	8	ADR70386	Adr70386 Poliovir
43	62	55.4	269	8	ADR28054	Adr28054 NPb poly
44	62	55.4	269	8	ADS17525	Ads17525 Amino ac
45	62	55.4	273	8	AD025157	Ado25157 Melanoma

ALIGNMENTS

RESULT 1					
ID	ABP60362	standard; peptide; 24 AA.			
AC	XX				
XX	ABP60362;				
DT	28-MAR-2003	(first entry)			
XX	XX				
DE	Streptavidin binding peptide SEQ ID NO 3.				
XX	XX				
KW	Streptavidin; protein chip; microtitre plate; detection.				
XX	XX				
OS	Synthetic.				
XX	XX				
FH	Key	Location/Qualifiers			
FT	Misc-difference 9..16				
FT		/label= unknown			
XX	XX				
PN	DE1013776-A1.				
XX	XX				
PD	02-OCT-2002.				
XX	XX				
PF	21-MAR-2001; 2001DE-01013776.				
XX	XX				
PR	21-MAR-2001; 2001DE-01013776.				
XX	XX				
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.				
XX	XX				
PI	Schmidt T;				
XX	XX				
DR	WPI; 2003-031166/03.				
XX	XX				
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.				
PT	modules.				
PS	Disclosure; Page 4; 18pp; German.				
XX	XX				
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (II), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily				


```

RESULT 4
ADOL9051
ID ADOL9051 standard; protein; 236 AA.
AC
XX ADOL9051;
XX
DT 26-AUG-2004 (first entry)
DE Murine antibody M1 Fab fragment heavy chain.
XX
KM Mouse; antibody; M1 Fab; heavy chain; Fab; scFv; abused drug; morphine;
KW THC; amphetamine; environmental hazard; toxic compound;
KV microbial process; metabolic process; drug monitoring;
XX pharmacological research.
OS
MS Mus sp.
XX
PN MO2004046733-A1.
XX
PP 03-JUN-2004.
PR 17-NOV-2003; 2003WO-FI000875.
PR 18-NOV-2002; 2002FI-00002048.
XX
PA (VALM ) VALTION TEKNILINEN TUTKIMUSKESKUS.
XX
PI Pulli T, Hoeyhyae M, Takkinen K, Soederlund H;
XX
DR WPI; 2004-420710/39.
XX
PT Non-competitive immunoassay for small analyte, useful for assaying drug
PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
PT reagent pair comprising a first binding partner and a second binding
PT partner.
XX
PS Claim 21; SEQ ID NO 2; 35pp; English.
XX
CC The invention relates to a non-competitive immunoassay for a small
CC analyte, comprising reacting a sample containing the analyte with a
CC reagent pair comprising a first binding partner that binds to the analyte
CC and a second binding partner that binds to the complex of the analyte and
CC the first binding partner, and determining the binding of the second
CC binding partner, thus indicating the presence of the analyte in the
CC sample. The first and second binding partners are antibody fragments Fab
CC or scFv. The reagent pair is useful in a non-competitive immunoassay for
CC a small analyte, particularly for assaying drugs of abuse e.g., morphine,
CC THC or amphetamine. The immunoassay is useful for detecting environmental
CC hazards, toxic compounds in food and feed, chemicals indicative of
CC ongoing processes (e.g., microbial processes in buildings, metabolic
CC processes of living organisms) and in clinical tests, drug monitoring and
CC pharmacological research. This sequence represents the murine antibody M1
CC Fab fragment heavy chain, used in the method of the invention.
XX
SQ Sequence 236 AA;
XX
Query Match 64.3%; Score 72; DB 8; Length 236;
Best Local Similarity 37.9%; Pred. No. 1.1;
Matches 11; Conservative 13; Mismatches 1; Indels 4; Gaps 1
QY 2 SHP-----QEKKXXXXXXXXXSWSHOFEK 26
DB 208 AHPASSTVKDKIVRDCGTSWSHOFEK 236

```

DT	26-AUG-2004	(first entry)
XX		
DE	Murine antibody M2 Fab fragment heavy chain.	
KW	Mouse; antibody; M2 Fab; heavy chain; Fab; scFv; abused drug; morphine;	
KW	THC; amphetamine; environmental hazard; toxic compound;	
KW	microbial process; metabolic process; drug monitoring;	
KW	pharmacological research.	
OS	Mus sp.	
XX		
PN	MO2004046733-A1.	
PD		
XX	03-JUN-2004.	
XX		
PF	17-NOV-2003; 2003WO-FI000875.	
XX		
PR	18-NOV-2002; 2002FI-00002048.	
PA	(VALM) VALTION TEKNIILLINEN TUTKIMUSKESKUS.	
PI	Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H,	
XX		
DR	WPI; 2004-420710/39.	
XX		
PT	Non-competitive immunoassay for small analyte, useful for assaying drug	
PT	of abuse (e.g., morphine), comprises reacting a sample of analyte with a	
PT	reagent pair comprising a first binding partner and a second binding	
PT	partner.	
XX		
PS	Claim 21; SEQ ID NO 4; 35pp; English.	
XX		
CC	The invention relates to a non-competitive immunoassay for a small	
CC	analyte, comprising reacting a sample containing the analyte with a	
CC	reagent pair comprising a first binding partner that binds to the analyte	
CC	and a second binding partner that binds to the complex of the analyte and	
CC	the first binding partner, and determining the binding of the second	
CC	binding partner, thus indicating the presence of the analyte in the	
CC	sample. The first and second binding partners are antibody fragments Fab	
CC	or scFv. The reagent pair is useful in a non-competitive immunoassay for	
CC	a small analyte, particularly for assaying drugs of abuse e.g., morphine,	
CC	THC or amphetamine. The immunoassay is useful for detecting environmental	
CC	hazard, toxic compounds in food and feed, chemicals indicative of	
CC	ongoing processes (e.g., microbial processes in buildings, metabolic	
CC	processes of living organisms) and in clinical tests, drug monitoring and	
CC	pharmacological research. This sequence represents the murine antibody M2	
CC	Fab fragment heavy chain, used in the method of the invention.	
XX		
SO	Sequence 236 AA;	
XX		
Query Match	64.3%; Score 72; DB 8; Length 236;	
Best Local Similarity	37.9%; Pred. No. 1.1;	
Matches	11; Conservative 13; Mismatches 1; Indels 4; Gaps 1,	
OY	2 SHP---OFEKXXXXXXXXXXWSHPQFX 26	
DB	208 AHPASSTVKKIVPRDCGTSWSPQFX 236	
XX		
RESULT 6		
ID	ABP60363 standard; peptide; 24 AA.	
XX		
AC	ABP60363;	
XX		
DT	28-MAR-2003 (first entry)	
DE	Streptavidin binding peptide SEQ ID NO 4.	
XX		
KW	Streptavidin; protein chip; microtitre plate; detection.	
XX		
OS	Synthetic.	
XX		

FH		Key	Location/Qualifiers
FT	Misc-difference	9..18	/label= unknown
FT	Misc-difference	22..24	/label= unknown
XX			
PN	DE01013776-A1.		
PD	02-OCT-2002.		
XX			
PF	21-MAR-2001; 200IDE-01013776.		
PR	21-MAR-2001; 200IDE-01013776.	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
PA	Schmidt T;		
P1	WPI; 2003-031166/03.		
XX			
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.		
PS	Disclosure; Page 4; 18pp; German.		
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Glu, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) so it may not be essential to remove it); facilitates detection and has easily controllable binding properties. (I) is particularly used for purifying FP from dilute solution in batch formats (which use sampler apparatus than column methods and result in lower loss of FP). The present sequence is that of a streptavidin binding peptide disclosed with the invention SQ Sequence 24 AA:		
Qy	Query Match : Best Local Similarity 62.5%; Score 70; Matches 21; Conservative 0; Mismatches 0; Indels 2; Gaps 1,		
Dd	1 WSHPOFEKKXXXXXXXXXWSHPQ 23 1 WSHPOFEKKXXXXXXXXX--HPQ 21		
RESULT 7			
ID	ADA00709 standard; protein; 209 AA.		
AC	ADA00709;		
DT	06-NOV-2003 (first entry)		
DE	Modified hNGAL related fusion protein SEQ ID NO:22.		
KM	mucin; human neutrophil gelatinase-associated lipocalin; hNGAL, rat alpha2-microglobulin-related protein; Azm; mouse 24p3/uterocalin; 24p3; mutagenesis; fusion protein; cytostatic; gene therapy; tumour imaging; cancer therapy. Synthetic. Homo sapiens.		
OS	Homo sapiens.		
XX			
FF	Key Location/Qualifiers		
FT	Peptide 1..21 /label= signal		
TF			

```

FT Protein 22..209 /note="modified hNGAL and Strep-tag II fusion protein"
FT FT /note= "22..199
FT Region /note="mature hNGAL"
FT FT 200..209
FT Region /note="Strep-tag II affinity tag"
PN MO2003029463-A2.
PD 10-APR-2003.
PF 18-SEP-2002; 2002WC-EP010490.
PX 27-SEP-2001; 2001WC-EP011213.
PR 16-APR-2002; 2002WC-EP004223.
PA (PIER-) PIERIS PROTEOLAB AG.
PI Skerra A, Schlehuber S;
PI WPI: 2003-372000/35.
PX DR N-PSDB; ADA00708.
DR
PT Generating a mutetin of a protein, e.g. human neutrophil gelatinase-
PT associated lipocalin, rat alpha2-microglobulin-related protein or mouse
PT 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
PX
PS Disclosure; Page 107-108; 122pp; English.
XX
XX The present invention describes a method for generating a mutetin of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the mutetin has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX mutetins of the protein. Also described: (1) a mutetin of hNGAL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the mutetin of
XX hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the mutetin; (3) a nucleic acid molecule
XX comprising a sequence encoding the mutetin of hNGAL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX the mutetin of hNGAL, A2m or 24p3 or the fusion protein described above,
XX and a carrier. The mutetins have cytototoxic activity, and can be used in
XX gene therapy. The method is useful in generating or producing a mutetin of
XX hNGAL, A2m or 24p3 or a fusion protein. The mutetin of hNGAL, A2m or 24p3
XX or the fusion protein is useful in detecting a given target by contacting
XX the mutetin with a sample suspected of containing the given target under
XX conditions that allow complex formation between the mutetin and the given
XX target, and determining the complexed mutetin by a suitable signal. The
XX given target is a protein or protein domain, a peptide, a nucleic acid
XX molecule, an organic molecule or a metal complex and the detection is
XX carried out for validation of the protein as a pharmacological drug
XX target. The mutetin may also be used in medicine, e.g. for tumour imaging
XX or directly for cancer therapy. The present sequence represents a
XX modified hNGAL and Strep-tag II fusion protein given in the
XX exemplification of the present invention.
SQ Sequence 209 AA;
Query Match 60.7%; Score 68; DB 6; Length 209;
Best Local Similarity 40.0%; Pred. NO. 3.4;
Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0
QY 2 SHPQFEKXXXXXXXXXWSPHPEK 26
DB 185 NHIVFPVIDCICIDGSAMSHPEK 209

```

```

XX AC ADA27293;
XX DT 20-NOV-2003 (first entry)
XX DE plasmid pHNGAL7 fragment protein.
XX DE Mutein; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;
XX KM alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
XX OS Synthetic.
XX FH Key
XX FH Peptide 1. .21
XX FT /note= "Ommpa signal sequence"
XX FT Protein 22. .258
XX FT /note= "Mature protein; fusion protein of hNGAL, Strep-
XX FT Tag II and albumin binding domain"
XX FT Protein 22. .199
XX FT /note= "Mature hNGAL"
XX FT Peptide 199. .209
XX FT /note= "Strep-Tag II affinity tag"
XX FT Protein 210. .258
XX FT /note= "Albumin binding domain of Protein G"
XX PN WO200302946d-A1.
XX PD 10-APR-2003
XX PF 27-SEP-2001; 2001WO-EP011213.
XX PR 27-SEP-2001; 2001WO-EP011213.
XX PA (PIER-) PIERIS PROTEOLAB AG.
XX PI Skerra A, Schlehuber S;
XX PI MPI; 2003-381639/36.
XX DR N-PSDB; ADA27287.
XX DT
XX PT Generating a mutein of a protein for validating the protein as drug
XX PT target by subjecting the protein to mutagenesis at sequence positions
XX PT corresponding to sequence positions of the human neutrophil gelatinase-
XX PT associated lipocalin.
XX PS Disclosure; Page 59-61; 68pp; English.
XX CC The present invention relates to a method for generating muteins of human
XX CC neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX CC microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX CC subjecting the protein to mutagenesis. In the method of generating a
XX CC mutein, a nucleic acid coding for the mutein of the protein resulting
XX CC from mutagenesis is operably fused at the 3' end with a gene coding for
XX CC the coat protein pill of a filamentous bacteriophage of the M13-family or
XX CC for a fragment of the coat protein. The present sequence is the protein
XX CC encoded by a fragment of plasmid pHNGAL7, used to illustrate the
XX CC invention. This sequence comprises human hNGAL, a Strep-tag II and an
XX CC albumin binding domain of Protein G.
XX SQ Sequence 258 AA;

Query Match 60.7%; Score 68; DB 6; Length 258;
Best Local Similarity 40.0%; Pred. No. 4.5;
Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Oy 2 SHPOFEKXXXXXXXXXSHPOFEK 26
: | | :|:|:|:|:|:|:|:|:|:|
Db 185 NHIVFPVPIDGCDISGSAKSHPOFEK 209

RESULT 9
ADA00702
ID ADA00702 standard; protein; 258 AA.

```

```

XX AC ADA00702;
XX DT 06-NOV-2003 (first entry)
XX DE Modified hNGAL related fusion protein SEQ ID NO:15.
XX DE Mutein; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX KM rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX KM 24p3; mutagenesis; fusion protein; cytosolic; gene therapy;
XX KM tumour imaging; cancer therapy.
XX OS Synthetic.
XX FH Homo sapiens.
XX FH Key
XX FH Peptide 1. .21
XX FT /label= signal
XX FT Protein 22. .258
XX FT /note= "modified hNGAL, Strep-tag II and albumin binding
XX FT domain of Protein G fusion protein"
XX FT Region 22. .199
XX FT /note= "Mature hNGAL"
XX FT Region 200. .209
XX FT /note= "Strep-tag II affinity tag"
XX FT Region 210. .258
XX FT /note= "albumin binding domain of Protein G"
XX PN WO2003029463-A2.
XX PD 10-APR-2003.
XX PF 18-SEP-2002; 2002WO-EP010490.
XX PR 27-SEP-2001; 2001WO-EP011213.
XX PR 16-APR-2002; 2002WO-EP004223.
XX PA (PIER-) PIERIS PROTEOLAB AG.
XX PI Skerra A, Schlehuber S;
XX PI MPI; 2003-372000/35.
XX DR N-PSDB; ADA00733.
XX DT
XX PT Generating a mutein of a protein, e.g. human neutrophil gelatinase-
XX PT associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX PT 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX PS Disclosure; Page 98-99; 122pp; English.
XX CC The present invention describes a method for generating a mutein of a
XX CC protein selected from a human neutrophil gelatinase-associated lipocalin
XX CC (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX CC 24p3/uterocalin (24p3), where the mutein has a detectable affinity to a
XX CC given target, comprising subjecting the protein to mutagenesis at
XX CC positions 93-94, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX CC muteins of the protein. Also described: (1) a mutein of hNGAL, A2m or
XX CC 24p3 having detectable binding affinity to a given target, obtained by
XX CC the method described above; (2) a fusion protein comprising the mutein of
XX CC hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX CC peptide, a signal sequence and/or an affinity tag is operably fused to
XX CC the amino or carboxy terminus of the mutein; (3) a nucleic acid molecule
XX CC comprising a sequence encoding the mutein of hNGAL, A2m or 24p3 or the
XX CC fusion protein of (2); and (4) a pharmaceutical composition comprising
XX CC the mutein of hNGAL, A2m or 24p3 or the fusion protein described above,
XX CC and a carrier. The muteins have cytostatic activity, and can be used in
XX CC gene therapy. The method is useful in generating or producing a mutein of
XX CC hNGAL, A2m or 24p3 or a fusion protein. The mutein of hNGAL, A2m or 24p3
XX CC or the fusion protein is useful in detecting a given target by contacting
XX CC the mutein with a sample suspected of containing the given target under
XX CC conditions that allow complex formation between the mutein and the given
XX CC target, and determining the complexed mutein by a suitable signal. The
XX CC given target is a protein or protein domain, a peptide, a nucleic acid

```


[illegible]

```

FT      Misc-difference 210
FT      /note= "amber stop codon"
FT      211. .400
FT      Region
FT      /note= "coat protein pIII fragment 217-406"
FT
FT      MO2003029463-AA2.
FT
FT      10-APR-2003.
FT
FT      18-SEP-2002; 2002MO-EP010490.
FT
FT      27-SEP-2001; 2001MO-EP011213.
FT      16-APR-2002; 2002MO-EP004223.
FT
FT      (PIER-) PIERIS PROTEOLAB AG.
FT
FT      Skerra A, Schlehuber S;
FT
FT      WPI; 2003-372000/35.
FT      DR
FT      N-PSDB; ADA00706.
FT
FT      Generating a mutetin of a protein, e.g. human neutrophil gelatinase-
FT      associated lipocalin, rat alpha2-microglobulin-related protein or mouse
FT      24p3/uterocalin, comprises subjecting the protein to mutagenesis.
FT
FT      Disclosure; Page 104-105; 122pp; English.
FT
FT      The present invention describes a method for generating a mutetin of a
FT      protein selected from a human neutrophil gelatinase-associated lipocalin
FT      (hNGL), rat alpha2-microglobulin-related protein (A2m) and a mouse
FT      24p3/uterocalin (24p3), where the mutetin has a detectable affinity to a
FT      given target, comprising subjecting the protein to mutagenesis at
FT      positions 33-54, 66-83, 94-106 and/or 123-136 of hNGL, resulting in
FT      mutetin of the protein. Also described: (1) a mutetin of hNGL, A2m or
FT      24p3 having detectable binding affinity to a given target, obtained by
FT      the method described above; (2) a fusion protein comprising the mutetin of
FT      hNGL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
FT      peptide, a signal sequence and/or an affinity tag is operably fused to
FT      the amino or carboxy terminus of the mutetin; (3) a nucleic acid molecule
FT      comprising a sequence encoding the mutetin of hNGL, A2m or 24p3 or the
FT      fusion protein of (2); and (4) a pharmaceutical composition comprising
FT      the mutetin of hNGL, A2m or 24p3 or the fusion protein described above,
FT      and a carrier. The mutetins have cytosratic activity, and can be used in
FT      gene therapy. The method is useful in generating or producing a mutetin of
FT      hNGL, A2m or 24p3 or a fusion protein. The mutetin of hNGL, A2m or 24p3
FT      or the fusion protein is useful in detecting a given target by contacting
FT      the mutetin with a sample suspected of containing the given target under
FT      conditions that allow complex formation between the mutetin and the given
FT      target, and determining the complexed mutetin by a suitable signal. The
FT      given target is a protein or protein domain, a peptide, a nucleic acid
FT      molecule, an organic molecule or a metal complex and the detection is
FT      carried out for validation of the protein as a pharmacological drug
FT      target. The mutetin may also be used in medicine, e.g. for tumour imaging
FT      or directly for cancer therapy. The present sequence represents a
FT      modified hNGL, Strep-tag II and phage coat protein pIII fragment fusion
FT      protein given in the exemplification of the present invention.
FT
SQ      Sequence 400 AA;
SQ
Qy      Query Match      60.7%; Score 68; DB 6; Length 400;
Db      Best Local Similarity 40.0%; Pred. No. 8.4;
Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0
Qy      2 SHPOFEKXXXXXXXXXXWSHPOFEK 26
Db      :|||:|||||:|||||:|||||:
Db      :|||:|||||:|||||:|||||:
Db      185 NHIVFPVIDICIDISAMSHPOFEK 209
Qy      RESULT 14
Db      ADA00701
Qy      ADA00701 standard; protein; 400 AA.
Db      ADA00701;

```

XX	06-NOV-2003	(first entry)	
DT	Modified hNGAL related fusion protein	SEQ ID NO:14.	
XX			
XX			
DE	mutin; human neutrophil gelatinase-associated lipocalin; hNGAL.		
XX	rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;		
KM	24p3; mutagenesis; fusion protein; cytostatic; gene therapy;		
KW	tumour imaging; cancer therapy.		
XX			
XX	Synthetic.		
OS	Homo sapiens.		
FT			
FH	Key	Location/Qualifiers	
FT	Peptide	1..21	
FT	Protein	/label= signal	
FT		22..400	
FT		/note= "modified hNGAL, Strep-tag II and phage coat	
FT		protein pIII fragment fusion protein"	
FT	Region	22..199	
FT		/note= "mature hNGAL"	
FT	Region	200..209	
FT		/note= "Strep-tag II affinity tag"	
FT	Misc-difference	210	
FT		/note= "amber stop codon"	
FT	Region	211..400	
FT		/note= "coat protein pIII fragment 217-406"	
XX			
PN	MO2003029463-A2.		
PD			
XX	10-APR-2003.		
XX			
PF	18-SEP-2002; 2002MO-EP010490.		
XX			
XX			
PR	27-SEP-2001; 2001MO-EP011213.		
PR	16-APR-2002; 2002MO-EP004223.		
XX			
PA	(PIER-) PIERIS PROTEOLAB AG.		
XX			
PI	Skerer A, Schlehuber S;		
XX			
DR	WPI; 2003-372000/35.		
XX	DR N-PSDB; ADA00732.		
PT	Generating a mutin of a protein, e.g. human neutrophil gelatinase-		
PT	associated lipocalin, rat alpha2-microglobulin-related protein or mouse		
PT	24p3/uterocalin, comprises subjecting the protein to mutagenesis.		
XX			
PS	Disclosure; Page 96-98; 122pp; English.		
XX			
XX			
CC	The present invention describes a method for generating a mutin of a		
CC	protein selected from a human neutrophil gelatinase-associated lipocalin		
CC	(hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse		
CC	24p3/uterocalin (24p3), where the mutin has a detectable affinity to a		
CC	given target, comprising subjecting the protein to mutagenesis at		
CC	positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in		
CC	mutins of the protein. Also described: (1) a mutin of hNGAL, A2m or		
CC	24p3 having detectable binding affinity to a given target, obtained by		
CC	the method described above; (2) a fusion protein comprising the mutin of		
CC	hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a		
CC	peptide, a signal sequence and/or an affinity tag is operably fused to		
CC	the amino or carboxy terminus of the mutin; (3) a nucleic acid molecule		
CC	comprising a sequence encoding the mutin of hNGAL, A2m or 24p3 or the		
CC	fusion protein of (2); and (4) a pharmaceutical composition comprising		
CC	the mutin of hNGAL, A2m or 24p3 or the fusion protein described above,		
CC	and a carrier. The mutins have cytostatic activity, and can be used in		
CC	gene therapy. The method is useful in generating or producing a mutin of		
CC	hNGAL, A2m or 24p3 or a fusion protein. The mutin of hNGAL, A2m or 24p3		
CC	or the fusion protein is useful in detecting a given target by contacting		
CC	conditions that allow complex formation between the mutin and the given		
CC	target, and determining the complexed mutin by a suitable signal. The		
CC	given target is a protein or protein domain, a peptide, a nucleic acid		

CC molecule, an organic molecule or a metal complex and the detection is
 CC carried out for validation of the protein as a pharmacological drug
 CC target. The molecule may also be used in medicine, e.g. for tumour imaging
 CC or directly for cancer therapy. The present sequence represents a
 CC modified hNGAL, Strep-tag II and phase coat protein pIII fragment fusion
 CC protein given in the exemplification of the present invention.
 XX

SQ Sequence 400 AA;

Query Match 60.7%; Score 68; DB 6; Length 400;

Best Local Similarity 40.0%; Pred. No. 8.4;
 Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SHPOEKXXXXXXXXXXSHPOFEK 26

Db 185 NHIVPPVPIDQIDGSAMSHPOFEK 209

RESULT 15

ADS20251
 ID ADS20251 standard; protein, 646 AA.

AC ADS20251;

DT 18-NOV-2004 (first entry)

DE Human aggrecanase modified ADAMTS4 (mTS4) protein - SEQ ID 49.

XX ADAMTS4, a disintegrin-like and metalloprotease;

KW thrombospondin type 1 motif 4; repolyisin; zinc metalloprotease;
 KM aggrecanase; osteopontin; antiinflammatory; antirheumatic;
 KM cystostatic; osteoarthritis; glioma; cancer; inflammatory joint;
 KM rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
 KM human; chromosome 1q21-q23; enzyme.

XX

XX Homo sapiens.

OS Synthetic.

XX WO2004011637-A2.

XX 05-FEB-2004.

XX 29-JUL-2003; 2003WO-US023484.

XX 29-JUL-2002; 2002US-0398721P.

XX (AMHP) WYETH.

XX (CORC/) CORCORAN C J.

XX (FLAN/) FLANNERY C R.

XX (ZENG/) ZENG W.

XX (RACI/) RACIE L A.

XX (MCDO/) MCDONAGH T.

XX (FREE/) FREEMAN B A.

XX (GEOR/) GEORGIADIS K E.

XX (LAVA/) LAVALLIE E R.

XX Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T,
 PI Freeman BA, Georgiadis KE, Lavallie ER;

XX WPI; 2004-143660/14.

XX

XX

XX

XX

XX

CC metalloproteases and include aggrecanases amongst their members. The
 CC protein of the invention demonstrates osteopontin, antiinflammatory,
 CC antiarthritic, antirheumatic and cytostatic activities and may be useful
 CC for treating aggrecanase-associated conditions, including osteoarthritis,
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
 CC arthritis, periodontal disease and Crohn's disease. The current sequence
 CC is that of a human aggrecanase modified ADAMTS4 (mTS4) protein of the
 CC invention.
 XX

SQ Sequence 646 AA;

Query Match 58.9%; Score 66; DB 8; Length 646;

Best Local Similarity 31.2%; Pred. No. 31;
 Matches 10; Conservative 12; Mismatches 4; Indels 6; Gaps 1;

Qy 1 WSHPO-----PEKXXXXXXXXXXSHPOFEK 26

Db 615 WLHRRQILILEILRRPFWAGRKGSAMSHPOFEK 646

Search completed: March 2, 2005, 13:02:50

Job time : 118.106 secs

XX The invention relates to a novel isolated, modified ADAMTS4 (a
 CC disintegrin-like and metalloprotease (repolysin type) with
 CC thrombospondin type 1 motif 4) protein with improved stability compared
 CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
 CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
 CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 80.8537 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ14
Perfect score: 112
Sequence: 1 whpqtfxkxxxxxxxxxwhpqtfxk 26

Scoring table: BLOSUM62GX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues
Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	88.4	24	14	US-10-026-578B-3
2	97	86.6	36	14	US-10-026-578B-10
3	84	75.0	36	14	US-10-026-578B-11
4	70	62.5	24	14	US-10-026-578B-4
5	66	58.9	646	16	US-10-628-432-49
6	66	58.9	858	16	US-10-628-432-27
7	64.5	57.6	485	16	US-10-628-432-47
8	64.5	57.6	697	16	US-10-628-432-24
9	64	57.1	24	9	US-09-809-517A-31
10	64	57.1	25	9	US-09-809-517A-34
11	64	57.1	245	17	US-10-887-228A-9
12	64	57.1	246	17	US-10-887-228A-9
13	64	57.1	252	17	US-10-887-228A-5

14	64	57.1	699	16	US-10-408-765A-434	Sequence 434, App
15	64	57.1	699	17	US-10-741-600-851	Sequence 851, App
16	64	57.1	699	17	US-10-741-600-852	Sequence 852, App
17	62	55.4	36	11	US-09-833-245-1103	Sequence 1103, App
18	62	55.4	56	11	US-09-833-245-1105	Sequence 1105, App
19	62	55.4	633	16	US-10-628-432-53	Sequence 53, App
20	61	54.5	21	9	US-09-809-517A-30	Sequence 30, App
21	61	54.5	22	9	US-09-809-517A-33	Sequence 33, App
22	61	54.5	46	15	US-10-432-262-12	Sequence 12, App
23	61	54.5	46	15	US-10-448-609-8	Sequence 8, App
24	61	54.5	117	10	US-09-977-137A-4	Sequence 4, App
25	61	54.5	117	10	US-09-977-137A-5	Sequence 5, App
26	61	54.5	117	10	US-09-977-137A-7	Sequence 7, App
27	61	54.5	117	10	US-09-977-137A-8	Sequence 8, App
28	61	54.5	117	10	US-09-977-137A-9	Sequence 9, App
29	61	54.5	117	10	US-09-977-137A-10	Sequence 10, App
30	61	54.5	117	10	US-09-977-137A-11	Sequence 11, App
31	61	54.5	117	10	US-09-977-137A-12	Sequence 12, App
32	61	54.5	118	10	US-09-977-137A-6	Sequence 6, App
33	61	54.5	661	15	US-10-354-983-8	Sequence 8, App
34	61	54.5	763	15	US-10-358-283-15	Sequence 15, App
35	61	54.5	845	16	US-10-628-432-40	Sequence 40, App
36	60	53.6	513	15	US-10-272-196-39	Sequence 39, App
37	59	52.7	771	15	US-10-369-493-16969	Sequence 16969, A
38	58	51.8	199	16	US-10-437-963-164100	Sequence 164100, A
39	58	51.8	622	15	US-10-425-114-51780	Sequence 51780, A
40	58	51.3	134	16	US-10-437-963-104065	Sequence 104065, A
41	57.5	51.3	158	15	US-10-424-599-184769	Sequence 184769, A
42	57.5	51.3	212	15	US-10-424-599-184075	Sequence 184075, A
43	57.5	51.3	252	16	US-10-437-963-200150	Sequence 200150, A
44	57.5	51.3	252	16	US-10-264-237-2199	Sequence 2199, App
45	57	50.9	121	15	US-10-264-237-2199	Sequence 2199, App

ALIGNMENTS

RESULT 1
US-10-026-578B-3
Sequence 3, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (10)..(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:


```

FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (10)-(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (11)-(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (12)-(12)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (13)-(13)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (14)-(14)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (15)-(15)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (16)-(16)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (17)-(17)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (18)-(18)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (22)-(22)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (23)-(23)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: MISC FEATURE
LOCATION: (24)-(24)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE: NAME/KEY: misc_feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-4

Query Match      62.5%; Score 70; DB 14; Length 24;
Best Local Similarity 91.3%; Pred. No. 0.14;
Matches 21; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

OY      1 WSHPOFEKXXXXXXXXXXWSHPQ 23
Db       1 WSHPOFEKXXXXXXXXXX--HPQ 21

RESULT 5
US-10-628-432-49
Sequence 49, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53

```

[illegible]

Db 456 SKKKFKCMVCGGSGSGSAMSHPOPEK 485

RESULT 8

US-10-628-432-24

Sequence 24, Application US/10628432

Publication No. US20040142863A1

GENERAL INFORMATION:

APPLICANT: Wyeth

TITLE OF INVENTION: Modified ADAMTS4 molecules

FILE REFERENCE: AM101378

CURRENT APPLICATION NUMBER: US/10/628,432

CURRENT FILING DATE: 2003-07-29

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn version 3.1

SEQ ID NO 24

LENGTH: 697

TYPE: PRT

ORGANISM: Artificial

FEATURE:

OTHER INFORMATION: Truncated ADAMTS4 molecule

US-10-628-432-24

Query Match 57.1%; Score 64; DB 16; Length 697;

Best Local Similarity 41.7%; Pred. No. 50;

Matches 11; Conservative 12; Mismatches 2; Indels 5; Gaps 1;

QY 2 SHPOPEK-----XXXXXXXXXXWSHPQPEK 26

Db 668 SKKKFKCMVCGGSGSGSAMSHPOPEK 697

RESULT 9

US-09-809-517A-31

Sequence 31, Application US/09809517A

Patent No. US20020034733A1

GENERAL INFORMATION:

APPLICANT: Lohmning, Corinna

TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote

TITLE OF INVENTION: Particles via disulfide bonds

FILE REFERENCE: MORPHO/11

CURRENT APPLICATION NUMBER: US/09/809,517A

CURRENT FILING DATE: 2001-03-15

PRIOR APPLICATION NUMBER: EP 99114072.4

PRIOR FILING DATE: 1999-07-20

PRIOR APPLICATION NUMBER: EP 00103551.8

PRIOR FILING DATE: 2000-02-18

NUMBER OF SEQ ID NOS: 41

SOFTWARE: PatentIn version 3.0

SEQ ID NO 31

LENGTH: 24

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-31

Query Match 57.1%; Score 64; DB 9; Length 24;

Best Local Similarity 41.7%; Pred. No. 0.82;

Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

QY 5 QPEK--XXXXXXXXXXWSHPQPEK 26

Db 1 EPEQKLISEEDLNGAPWSHPQPEK 24

RESULT 10

US-09-809-517A-34

Sequence 34, Application US/09809517A

Patent No. US20020034733A1

GENERAL INFORMATION:

APPLICANT: Lohmning, Corinna

TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/proce

TITLE OF INVENTION: particles via disulfide bonds

FILE REFERENCE: MORPHO/11

CURRENT APPLICATION NUMBER: US/09/809,517A

CURRENT FILING DATE: 2001-03-15

PRIOR APPLICATION NUMBER: EP 99114072.4

PRIOR FILING DATE: 1999-07-20

PRIOR APPLICATION NUMBER: EP 00103551.8

PRIOR FILING DATE: 2000-02-18

NUMBER OF SEQ ID NOS: 41

SOFTWARE: PatentIn version 3.0

SEQ ID NO 34

LENGTH: 25

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-34

Query Match 57.1%; Score 64; DB 9; Length 25;

Best Local Similarity 41.7%; Pred. No. 0.86;

Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

QY 5 QPEK--XXXXXXXXXXWSHPQPEK 26

Db 2 EPEQKLISEEDLNGAPWSHPQPEK 25

RESULT 11

US-10-887-228A-1

Sequence 1, Application US/10887228A

Publication No. US20050037402A1

GENERAL INFORMATION:

APPLICANT: Schering AG

TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio

TITLE OF INVENTION: Substance That is Contained in a Sample

FILE REFERENCE: S30569US

CURRENT APPLICATION NUMBER: US/10/887,228A

CURRENT FILING DATE: 2004-07-09

PRIOR APPLICATION NUMBER: DE 103 31 093.2

PRIOR FILING DATE: 2003-07-09

PRIOR APPLICATION NUMBER: US 60/478,262

PRIOR FILING DATE: 2003-07-16

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 245

TYPE: PRT

ORGANISM: artificial

FEATURE:

OTHER INFORMATION: Fab fragment of variable heavy chain

US-10-887-228A-1

Query Match 57.1%; Score 64; DB 17; Length 245;

Best Local Similarity 41.7%; Pred. No. 16;

Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

QY 5 QPEK--XXXXXXXXXXWSHPQPEK 26

Db 222 EPEQKLISEEDLNGAPWSHPQPEK 245

RESULT 12

US-10-887-228A-9

Sequence 9, Application US/10887228A

Publication No. US20050037402A1

GENERAL INFORMATION:

APPLICANT: Schering AG

TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio

TITLE OF INVENTION: Substance That is Contained in a Sample

FILE REFERENCE: S30569US

CURRENT APPLICATION NUMBER: US/10/887,228A

CURRENT FILING DATE: 2004-07-09

PRIOR APPLICATION NUMBER: DE 103 31 093.2

```
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ PRIOR FILING DATE: 2003-07-16
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 9
/ LENGTH: 246
/ TYPE: PR
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match          57.1%; Score 64; DB 17; Length 246;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

Qy 5 QPEK-XXXXXXXXXXSHPOPEK 26
Db 223 EPEQKLISEEDLNGAPWMSHPQPEK 246

RESULT 13
US-10-887-228A-5
/ Sequence 5, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ FILE REFERENCE: S30569US
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 5
/ LENGTH: 252
/ TYPE: PR
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-5

Query Match          57.1%; Score 64; DB 17; Length 252;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

Qy 5 QPEK-XXXXXXXXXXSHPOPEK 26
Db 229 EPEQKLISEEDLNGAPWMSHPQPEK 252

RESULT 14
US-10-408-765A-434
/ Sequence 434, Application US/10408765A
/ Publication No. US20040101874A1
/ GENERAL INFORMATION:
/ APPLICANT: Ghosh, Soumitra S.
/ APPLICANT: Rahy, Bin D.
/ APPLICANT: Zhang, Bing
/ APPLICANT: Gibson, Bradford W.
/ APPLICANT: Taylor, Steven W.
/ APPLICANT: Glenn, Gary M.
/ APPLICANT: Marnock, Dale E.
/ TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
/ FILE REFERENCE: 660088.465
/ CURRENT APPLICATION NUMBER: US/10/408,765A
/ CURRENT FILING DATE: 2003-04-04
```

```
/ NUMBER OF SEQ ID NOS: 3077
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 434
/ LENGTH: 699
/ TYPE: PR
/ ORGANISM: Homo sapiens
US-10-408-765A-434

Query Match          57.1%; Score 64; DB 16; Length 699;
Best Local Similarity 37.5%; Pred. No. 59;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HPQEKXXXXXXXXXXSHPOPEK 26
Db 67 HPDENKDVSTENGHHFMSHPDREK 90

RESULT 15
US-10-741-600-851
/ Sequence 851, Application US/10741600
/ Publication No. US20050026169A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CI001499
/ CURRENT APPLICATION NUMBER: US/10/741,600
/ PRIOR FILING DATE: 2003-12-22
/ NUMBER OF SEQ ID NOS: 73997
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 851
/ LENGTH: 699
/ TYPE: PR
/ ORGANISM: Homo sapiens
US-10-741-600-851

Query Match          57.1%; Score 64; DB 17; Length 699;
Best Local Similarity 37.5%; Pred. No. 59;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HPQEKXXXXXXXXXXSHPOPEK 26
Db 67 HPDENKDVSTENGHHFMSHPDREK 90

Search completed: March 2, 2005, 14:19:02
Job time : 80.8537 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - prodein search, using sw model

Run on: March 2, 2005, 12:20:44, Search time 29.5935 Seconds
(Without alignments)
65.585 Million cell updates/sec

Title: SEQ14
Perfect score: 112
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfek 26

Scoring table: BLOSUM62GX
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/ECTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	57.1	24	US-09-809-517A-31	Sequence 31, Appl
2	64	57.1	25	US-09-809-517A-34	Sequence 34, Appl
3	64	57.1	699	US-09-538-092-995	Sequence 995, App
4	61	54.5	21	US-09-809-517A-30	Sequence 30, Appl
5	61	54.5	22	US-09-809-517A-33	Sequence 33, Appl
6	61	54.5	117	US-09-977-137A-4	Sequence 4, Appl
7	61	54.5	117	US-09-977-137A-5	Sequence 5, Appl
8	61	54.5	117	US-09-977-137A-7	Sequence 7, Appl
9	61	54.5	117	US-09-977-137A-8	Sequence 8, Appl
10	61	54.5	117	US-09-977-137A-9	Sequence 9, Appl
11	61	54.5	117	US-09-977-137A-10	Sequence 10, Appl
12	61	54.5	117	US-09-977-137A-11	Sequence 11, Appl
13	61	54.5	117	US-09-977-137A-12	Sequence 12, Appl
14	61	54.5	118	US-09-977-137A-6	Sequence 6, Appl
15	57	50.9	660	US-08-770-761A-2	Sequence 2, Appl
16	57	50.9	662	US-08-770-761A-5	Sequence 5, Appl
17	57	50.9	705	US-08-770-761A-7	Sequence 7, Appl
18	55.5	49.6	394	US-09-248-796A-11857	Sequence 11857, A
19	53	47.3	10	US-09-809-517A-6	Sequence 6, Appl
20	53	47.3	241	US-09-328-352-6364	Sequence 6364, Ap
21	53	47.3	463	US-09-370-767-48444	Sequence 48444, A
22	52	46.4	244	US-09-003-287-6	Sequence 6, Appl
23	52	46.4	244	US-09-003-287-8	Sequence 8, Appl
24	52	46.4	244	US-09-518-988-2	Sequence 2, Appl
25	52	46.4	386	US-08-895-707-7	Sequence 7, Appl
26	52	46.4	816	US-09-949-016-10904	Sequence 10904, A
27	52	46.4	922	US-09-252-991A-32759	Sequence 32759, A

28	52	46.4	2123	4	US-09-949-016-7517	Sequence 7517, Ap
29	52	46.4	3324	4	US-09-902-540-9732	Sequence 9732, Ap
30	52	46.4	4070	4	US-09-961-403-7	Sequence 7, Appl
31	52	46.4	3088	4	US-09-562-702A-8	Sequence 8, Appl
32	52	46.4	3089	4	US-09-562-702A-4	Sequence 4, Appl
33	52	46.4	3110	4	US-09-562-702A-2	Sequence 2, Appl
34	52	46.4	3110	4	US-09-562-702A-6	Sequence 6, Appl
35	52	46.4	3110	4	US-09-561-709B-7	Sequence 7, Appl
36	52	46.4	3110	4	US-09-917-254-86	Sequence 86, Appl
37	52	46.4	3110	4	US-09-949-016-5937	Sequence 5937, Ap
38	52	46.4	3111	2	US-08-460-309-4	Sequence 4, Appl
39	52	46.4	3111	2	US-08-125-077-4	Sequence 2, Appl
40	51	45.5	8	3	US-08-948-097-2	Sequence 2, Appl
41	51	45.5	8	3	US-09-382-950-7	Sequence 7, Appl
42	51	45.5	8	3	US-09-382-736B-8	Sequence 8, Appl
43	51	45.5	8	4	US-09-619-103-9	Sequence 9, Appl
44	51	45.5	8	4	US-10-104-218-5	Sequence 5, Appl
45	51	45.5	8	4	US-09-809-517A-9	Sequence 9, Appl

ALIGNMENTS

```
RESULT 1
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on i
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31
Query Match          57.1% Score 64; DB 4; Length 24;
Best Local Similarity 41.7%; Pred. No. 0.073;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;
QY 5 QFEK-XXXXXXXXXWHPQFEK 26
   |||:|||||:|||||:
Db 1 EFEKLISEEDLNGAPWHPQFEK 24
RESULT 2
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on i
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match
Best Local Similarity 57.1%; Score 64; DB 4; Length 25;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

Qy 5 OFEK-XXXXXXXXXXWSHPQEK 26
: ||: :::::::::::|||||
Db 2 EPEQKLISEEDLNGAPWSHPQEK 25

RESULT 3
US-09-538-092-995
; Sequence 995, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 995
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P23327
US-09-538-092-995

Query Match
Best Local Similarity 57.1%; Score 64; DB 4; Length 699;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HPQEKXXXXXXXXXXWSHPQEK 26
: ||: :::::::::::|||||
Db 67 HPDENKDVSTENGHHFWSHPDREK 90

RESULT 4
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21

```

```

TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-517A-10

Query Match
Best Local Similarity 54.5%; Score 61; DB 4; Length 21;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 26
: : : : :
DB 4 YKDDDKGAPWSHPOFEK 21

RESULT 5
US-09-809-517A-33
Sequence 33, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corina
TITLE OF INVENTION: Particles via disulfide bonds
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 9914072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 33
LENGTH: 22
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match
Best Local Similarity 54.5%; Score 61; DB 4; Length 22;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 26
: : : : :
DB 5 YKDDDKGAPWSHPOFEK 22

RESULT 6
US-09-977-137A-4
Sequence 4, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Caguiat, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

```


[illegible]

```

RESULT 9
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 26
   :XXXXXXXXXXXXXXXXXXXX
DB 100 RKNVSCPSANSHPOFEK 117

RESULT 10
US-09-977-137A-9
; Sequence 9, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 26
   :XXXXXXXXXXXXXXXXXXXX
DB 100 RKNVSCPSANSHPOFEK 117

RESULT 11
US-09-977-137A-10
; Sequence 10, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:

```

```

; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

```

```

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches      8; Conservative 10; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 26
        :XXXXXXXXXXXXXXXXXXXX
DB      100 RKGWVSCPSSAMSHPOFEK 117

```

```

RESULT 12
US-09-977-137A-11
; Sequence 11, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

```

```

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches      8; Conservative 10; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 26
        :XXXXXXXXXXXXXXXXXXXX
DB      100 RKGWVSCPSSAMSHPOFEK 117

```

```

RESULT 13
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12

```

```

; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

```

```

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches      8; Conservative 10; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 26
        :XXXXXXXXXXXXXXXXXXXX
DB      100 RKGWVSCPSSAMSHPOFEK 117

```

```

RESULT 14
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

```

```

Query Match          54.5%; Score 61; DB 4; Length 118;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches      8; Conservative 10; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 26
        :XXXXXXXXXXXXXXXXXXXX
DB      101 RKGWVSCPSSAMSHPOFEK 118

```

```

RESULT 15
US-08-770-761A-2
; Sequence 2, Application US/08770761A
; Patent No. 5814503
; GENERAL INFORMATION:
; APPLICANT: Kovacevic, Steven
; APPLICANT: Otto, Keith A.
; APPLICANT: Rao, Ramachandra N.
; TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE
; TITLE OF INVENTION: REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:

```

```

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/770,761A
/ FILING DATE: 19-DEC-1996
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: GAYLO, Paul J.
/ REGISTRATION NUMBER: 36,808
/ REFERENCE/DOCKET NUMBER: X-10136
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 317-376-0756
/ TELEFAX: 317-277-1917
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 660 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULAR TYPE: peptide
/ US-08-770-761A-2

```

```

Query Match          50.9%; Score 57; DB 2; Length 660;
Best Local Similarity 31.8%; Pred. No. 80;
Matches 7; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY      3 HPQFEKXXXXXXXXXXWSHPQF 24
          | | | | | | | | | | | |
DB      637 HSYLHKDEGNPEGSAWRHPQF 658

```

Search completed: March 2, 2005, 12:25:38
Job time : 30.5935 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 20.6341 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ15
Perfect score: 113
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfex 27

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	54.0	433	2 H83444	probable cytochrome
2	59	52.2	219	2 C96510	hypothetical prote
3	58	51.3	103	2 E90973	hypothetical prote
4	58	51.3	103	2 A85821	unknown protein en
5	58	51.3	103	2 C85713	unknown protein en
6	58	51.3	103	2 A90904	hypothetical prote
7	57.5	50.9	699	2 A54660	histidine rich cal
8	56	49.6	242	2 T40523	hypothetical prote
9	55	48.7	207	2 JH0145	hypothetical 24.1k
10	55	48.7	224	2 S71749	DCI. protein precu
11	54.5	48.2	1087	2 T22847	hypothetical prote
12	54.5	48.2	1250	2 T22845	hypothetical prote
13	54	47.8	211	2 B71058	hypothetical prote
14	54	47.8	540	2 T10892	probable calnexin
15	54	47.8	600	2 I49281	feritin alpha pre
16	54	47.8	947	2 T12787	valyl-tRNA synthet
17	54	47.8	947	2 B97567	valyl-tRNA synthet
18	54	47.8	2037	2 S37178	fatty-acyl-CoA syn
19	53.5	47.3	304	2 A81069	mr restriction by
20	53.5	47.3	304	2 A40368	mr protein - Esch
21	53	46.9	367	2 F64502	probable GTP-Dindi
22	53	46.9	403	2 A12363	transposase allr46
23	53	46.9	403	2 AE2518	transposase allr732
24	53	46.9	421	2 G82422	anaerobic glycerol
25	53	46.9	421	2 A47713	chitin deacetylase
26	53	46.9	453	2 D95342	Nosd periplasmic c
27	53	46.9	478	2 T40362	conserved hypochet
28	53	46.9	684	2 S52835	hypothetical prote
29	53	46.9	812	1 A36477	fibroblast growth

30	53	46.9	814	1 A39752	fibroblast growth
31	53	46.9	909	2 A03419	phage host specific
32	53	46.9	1209	2 T31657	reverse transcript
33	53	46.9	2471	2 T42977	large tegument pro
34	52.5	46.5	342	2 S37596	protein kinase MCP
35	52	46.0	332	2 I39927	ABC transporter (b
36	52	46.0	334	2 AF1151	6-phosphogluconate
37	52	46.0	334	2 AF9136	probable 6-phospho
38	52	46.0	362	2 C90674	hypothetical prote
39	52	46.0	362	2 C64758	Yahn protein - Bsc
40	52	46.0	365	2 F85524	hypothetical prote
41	52	46.0	391	2 G86718	replication protei
42	52	46.0	403	2 AB1951	transposase alr115
43	52	46.0	449	2 S07714	T64 protein precu
44	52	46.0	450	2 E70590	3-phosphoshikimate
45	52	46.0	451	2 I50131	clusterin - quail

ALIGNMENTS

```

RESULT 1
H83444
probable cytochrome c PAL600 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H83444
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Larbig, K.; Lam,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: AB2950; PMID:20437337; PMID:10984043
A:Accession: H83444
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <STO-
A:Cross-references: UNIPROT:Q913C1; GB:AE004588; GB:AE004091; NID:99947563; PIDN:AMG049
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PAL600
C:Superfamily: membrane-bound alcohol dehydrogenase, cytochrome c subunit; cytochrome c
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:58/61/Binding site: heme (Cys) (covalent) #status predicted
F:62/Binding site: heme iron (His) (axial ligand) #status predicted
F:204/207/Binding site: heme (Cys) (covalent) #status predicted
F:208/Binding site: heme iron (His) (axial ligand) #status predicted
F:336/339/Binding site: heme (Cys) (covalent) #status predicted
F:340/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match          54.0%; Score 61; DB 2; Length 433;
Best Local Similarity 28.0%; Pred. No. 8.7;
Matches 7; Conservative 15; Mismatches 3; Indels 0; Gaps 0;

QY 1 WSHPOFKKXXXXXXXXXWSHPQF 25
DB 99 WSPAPFERMRHGVARDGSYLPAP 123

RESULT 2
C96510
hypothetical protein P2G19.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: C96510
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Comn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Luo, Z.A.; Luoro, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

```

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C96510
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-219 <STO>
 A:Cross-references: UNIPROT:Q9C642; GB:AE005173; NID:91121777; PIDN:AA634254.1; GSPDB:C
 C:Genetics:
 A:Gene: F2C19.1
 A:Map position: 1
 C:Superfamily: tomato DCL protein

Query Match 52.2%; Score 59; DB 2; Length 219;
 Best Local Similarity 33.3%; Pred. No. 7.5;
 Matches 8; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

QY 3 HPOFEKXXXXXXXXXXSHPOFEK 26
 ||:|||||:|||||:|||||
 Db 146 HPECEKKKGGCIGDIYIMVGHHPDFE 169

RESULT 3
 E90973
 Hypothetical protein EC62757 [imported] - *Escherichia coli* (strain O157:H7, substrain RI
 C:Species: *Escherichia coli*
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C:Accession: E90973
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: E90973
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <HAV>
 A:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BA936180.1; PID:913362225; GSPDB:C
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: EC62757

Query Match 51.3%; Score 58; DB 2; Length 103;
 Best Local Similarity 30.0%; Pred. No. 4.3;
 Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXXSHPOFEK 27
 ||:|||||:|||||:|||||
 Db 5 KEMPVVRNRYGCMTHPEYER 24

RESULT 4
 A85821
 Unknown protein encoded within prophage CP-933U [imported] - *Escherichia coli* (strain O1
 C:Species: *Escherichia coli*
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C:Accession: A85821
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouls, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: A85821
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <STO>
 A:Cross-references: UNIPROT:Q8X4V0; GB:AE005174; NID:912515136; PIDN:AA657029.1; GSPDB:C
 A:Experimental source: strain O157:H7, substrain EDU933
 C:Genetics:
 A:Gene: Z3120

Query Match 51.3%; Score 58; DB 2; Length 103;
 Best Local Similarity 30.0%; Pred. No. 4.3;

Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;
 QY 8 KXXXXXXXXXXSHPOFEK 27
 ||:|||||:|||||:|||||
 Db 5 KEMPVVRNRYGCMTHPEYER 24

RESULT 5
 C85713
 Unknown protein encoded within prophage CP-9330 [imported] - *Escherichia coli* (strain O1
 C:Species: *Escherichia coli*
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C:Accession: C85713
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouls, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: C85713
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <STO>
 A:Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:912515046; PIDN:AA656167.1; GSPDB:C
 A:Experimental source: strain O157:H7, substrain EDU933
 C:Genetics:
 A:Gene: Z2097

Query Match 51.3%; Score 58; DB 2; Length 103;
 Best Local Similarity 30.0%; Pred. No. 4.3;
 Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXXSHPOFEK 27
 ||:|||||:|||||:|||||
 Db 5 KEMPVVRDYGVTWTHPEYER 24

RESULT 6
 A90904
 Hypothetical protein EC62201 [imported] - *Escherichia coli* (strain O157:H7, substrain RI
 C:Species: *Escherichia coli*
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C:Accession: A90904
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: A90904
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <HAV>
 A:Cross-references: UNIPROT:Q8X549; GB:BA000007; PIDN:BA936167.1; PID:913361667; GSPDB:C
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: EC62201

Query Match 51.3%; Score 58; DB 2; Length 103;
 Best Local Similarity 30.0%; Pred. No. 4.3;
 Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXXSHPOFEK 27
 ||:|||||:|||||:|||||
 Db 5 KEMPVVRDYGVTWTHPEYER 24

RESULT 7
 A54660
 histidine rich calcium binding protein - human
 C:Species: *Homo sapiens* (man)
 C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 09-Jul-2004
 C:Accession: A54660
 R:Hofmann, S.L.; Topham, M.; Heien, C.L.; Francke, U.
 Genomics 9, 656-669, 1991

```

A>Title: cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum protein, and its
A:Reference number: A54660; MUID:91244309; PMID:2037293
A:Accession: A54660
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-699 <HOP>
A:Cross-references: UNIPROT:P23327; GB:M60052; NID:g183918; PIDN:AAA88071.1; PID:g183919
C:Genetics:
A:Gene: GDB:HRC
A:Cross-references: GDB:I26369; OMIM:142705
A:Map position: 19q13.3-19q13.3
C:Keywords: calcium binding

Query Match          50.9%; Score 57.5; DB 2; Length 699;
Best Local Similarity 36.0%; Pred. No. 47;
Matches      9; Conservative 11; Mismatches    4; Indels   1; Gaps   1;

Oy      3 HPQFEKXXXXXXXXXXWSHPQFEK 27
        |||.....|||
Db       67 HPD-ENKDVSTENGHHFMSHPDREK 90

RESULT 8
T40523
hypothetical protein SPBC530.07c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T40523
R:Layne, M.; Wood, V.; Rajandream, M.A.; Barrel, B.G.; Rieger, M.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21934
A:Accession: T40523
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-242 <LYN>
A:Cross-references: UNIPROT:O59743; EMBL:AL023634; PIDN:CAA19173.1; GSPDB:GN00067; SPDB:
A:Experimental source: strain 972h-; cosmid c530
C:Genetics:
A:Gene: SPDB:SPBC530.07c
A:Map position: 2

Query Match          49.6%; Score 56; DB 2; Length 242;
Best Local Similarity 30.4%; Pred. No. 22;
Matches      7; Conservative 12; Mismatches    4; Indels   0; Gaps   0;

Oy      5 QPEKXXXXXXXXXXWSHPQFEK 27
        |||.....|||
Db       180 QPKNCPSEMEPVKMGQPSFEK 202

RESULT 9
JH0145
hypothetical 24.1k protein - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C:Accession: JH0145
R:Schechtman, M.G.
Gene 88, 159-165, 1990
A>Title: Characterization of telomere DNA from Neurospora crassa.
A:Reference number: JH0145; MUID:90269603; PMID:1971801
A:Accession: JH0145
A:Molecule type: DNA
A:Residues: 1-207 <SCH>
A:Cross-references: UNIPROT:O01400; GB:M37064; NID:g168912; PIDN:AAC13656.1; PID:g529466
A:Experimental source: strain OR23-IV-A
A>Note: translation of the nucleotide sequence is not complete
C:Superfamily: Neurospora crassa hypothetical 24.1k protein

Query Match          48.7%; Score 55; DB 2; Length 207;
Best Local Similarity 30.0%; Pred. No. 26;
Matches      6; Conservative 13; Mismatches    1; Indels   0; Gaps   0;

Oy      6 FEKXXXXXXXXXXWSHPQF 25

```

```

Db      119 FEALITSHQYRSIMNHPRF 138
|||.....:|||:|
RESULT 10
S71749
DCL protein precursor, chloroplast - tomato
C:Species: Lycopersicon esculentum (tomato)
C:date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: S71749; EMBL:U55278; NID:G1323697; PID:G1323698
R:Keddie, J.S.; Carroll, B.; Jones, J.D.G.; Grisseem, W.
EMBO J. 15, 4208-4217, 1996
A>Title: The DCL gene of tomato is required for chloroplast development and palisade cell
A:Reference number: S71749; MOID:97015121; PMID:8861949
A:Accession: S71749
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-224 <KEP>-
A:Cross-references: UNIPROT:Q42463; EMBL:U55278; NID:G1323697; PID:G1323698
A:Experimental source: strain Moneymaker
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1996
A:Accession: S71748
A:Molecule type: mRNA
A:Residues: 1-224 <KEN>-
A:Cross-references: EMBL:U55219; NID:G1305530; PIDD:AAC49433.1; PID:G1305531
A:Experimental source: strain Moneymaker
C:Genetics:
A:Map position: 4
A:Genome: nuclear
A:Introns: 128/2; 166/3
C:function:
A>Description: required for both chloroplast development and palisade cell morphogenesis
C:Superfamily: tomato DCL protein
C:Keywords: chloroplast
F:1-50/Domain: transit peptide (chloroplast) #status predicted <TMP>
F:51-224/Product: DCL protein #status predicted <MAT>

Query Match          48.7%; Score 55; DB 2; Length 224;
Best Local Similarity 29.2%; Pred. No. 28;
Matches    7; Conservative   13; Mismatches     4; Indels    0; Gaps    0;

QY        3 HQPERXXXXXXXXXXMSHPQE 26
           ||::|::|::|::|::|::|::|::|
Db       151 HPECCKIGPGVDTVTGYHDPFE 174

RESULT 11
T22847
hypochlorite protein F57C7.lb - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22847
R:white, S.
Submitted to the EMBL Data Library, February 1996
A:Reference number: Z19625
A:Accession: T22847
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1087 <WIL>-
A:Cross-references: UNIPROT:Q20948; EMBL:Z69646; PIDD:CAA93475.1; GSPPDB:GN00028; CESP:F5
A:Experimental source: clone F57C7
C:Genetics:
A:Gene: CESP:F57C7.lb
A:Map position: X
A:Introns: 262/3; 351/3; 391/3; 627/3; 795/2; 957/1
F:307-364/Domain: bromodomain homology <BRO1>
F:579-636/Domain: bromodomain homology <BRO2>

Query Match          48.2%; Score 54.5; DB 2; Length 1087;
Best Local Similarity 22.9%; Pred. No. 21e+02;
Matches    8; Conservative   15; Mismatches     3; Indels    9; Gaps    1;

```

Db 524 AHQETPRKKNPTLIEWKHLVPRWQGIPEWQK 558

RESULT 12

T22845

hypothetical protein F57C7.1a - *Caenorhabditis elegans*C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C/Accession: T22845

R/White, S.

Submitted to the EMBL Data Library, February 1996

A/Reference number: Z19625

A/Accession: T22845

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 1-1250 <MTL>

A/Cross-references: UNIPROT:Q020947; EMBL:Z69646; PIDN:CAA93473.1; GSPDB:GN00028; CESP:FS

A/Experimental source: clone F57C7

C/Genetics:

A/Map position: X

A/Intons: 262/3; 351/3; 391/3; 431/3; 667/3; 835/2; 997/1; 1127/1; 1218/3

F/307-364/Domain: bromodomain homology <BRO1>

F/619-676/Domain: bromodomain homology <BRO2>

Query Match 48.2%; Score 54.5; DB 2; Length 1250;

Best Local Similarity 22.9%; Pred. No. 2.5e+02;

Matches 8; Conservative 15; Mismatches 3; Indels 9; Gaps 1;

QY 2 SHPQFEKXXXXXXXXXWSH-----PQFEK 27

Db 564 AHQETPRKKNPTLIEWKHLVPRWQGIPEWQK 598

RESULT 13

B71058

hypothetical protein PH160 - *Pyrococcus horikoshii*C/Species: *Pyrococcus horikoshii*

C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004

C/Accession: B71058

R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic e

A/Reference number: A71000; MUID:9834437; PMID:9679194

A/Accession: B71058

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-211 <KAM>

A/Cross-references: UNIPROT:Q58874; GB:AP000005; NID:93236132; PIDN:BAA30260.1; PID:G325

A/Experimental source: strain OT3

A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C/Genetics:

A/Genes: PH160

Query Match 47.8%; Score 54; DB 2; Length 211;

Best Local Similarity 31.0%; Pred. No. 36;

Matches 9; Conservative 14; Mismatches 4; Indels 2; Gaps 1;

QY 1 WSHPQFEKXXXXXXXXXWSH--HPQFEK 27

Db 158 WSHPQFEKXXXXXXXXXWSH--HPQFEK 186

RESULT 14

T10892

probable calnexin - *Jerusalem artichoke*C/Species: *Helianthus tuberosus* (*Jerusalem artichoke*)

C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C/Accession: T10892

R/Haeflitz, M.; Jeltsch, J.; Lesot, A.; Michalak, M.; Durst, F.

Submitted to the EMBL Data Library, July 1994

A/Description: Cloning and characterization of a cDNA encoding an analog of the calnexin

A/Reference number: Z17201

A/Accession: T10892

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-540 <HAS>

A/Cross-references: UNIPROT:Q39994; EMBL:Z35108; NID:9510906; PID:9510907

A/Experimental source: cv. blanc commun; tuber slice; parenchyma

C/Function:

A/Description: ER-bound chaperone

C/Superfamily: calnexin

C/Keywords: endoplasmic reticulum; molecular chaperone

Query Match 47.8%; Score 54; DB 2; Length 540;

Best Local Similarity 26.1%; Pred. No. 1.1e+02;

Matches 6; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPQFEKXXXXXXXXXWSHP 23

Db 301 WEAQIENPKESAPGCGEWRP 323

RESULT 15

I49281

fertilin alpha precursor - mouse (fragment)

C/Species: *Mus musculus* (house mouse)

C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Aug-2004

C/Accession: I49281

R/Wolfsberg, T.G.; Stralight, P.D.; Gereña, R.L.; Huovila, A.P.; Primakoff, P.; Wyles, D.

Dev. Biol. 169, 378-383, 1995

A/Title: ADAM, a widely distributed and developmentally regulated gene family encoding m

A/Reference number: I48100; MUID:95268891; PMID:7750654

A/Accession: I49281

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-600 <RES>

A/Cross-references: EMBL:U22056; NID:9965009; PIDN:AAA74920.1; PID:9965010

C/Genetics:

A/Genes: ADAM 1

C/Superfamily: disintegrin homology

F/246-326/Domain: disintegrin homology <DIS>

F/180/Active site: Glu #status predicted

Query Match 47.8%; Score 54; DB 2; Length 600;

Best Local Similarity 25.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 2 SHPQFEKXXXXXXXXXWSHPQF 25

Db 22 SHPQFEKXXXXXXXXXWSHPQF 45

Search completed: March 2, 2005, 12:29:01

Job time : 21.6341 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 100.317 Seconds
(without alignment)
137.824 Million cell updates/sec

Title: SEQ15
Perfect score: 113
Sequence: 1 whpbfekxxxxxxxxxxxxwhpbfek 27

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03: *
1: uniprot_gproc: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	60.2	107	2	062315
2	67	59.3	704	2	08CJD2
3	62	54.9	743	2	08CH90
4	61	54.0	433	2	0913C1
5	60	53.1	224	2	08RV8
6	59	52.2	219	2	09C642
7	59	52.2	389	2	09AD61
8	58.5	51.8	208	2	062FR3
9	58.5	51.8	256	2	0651B2
10	58	51.3	103	2	08X549
11	58	51.3	103	2	08X4V0
12	58	51.3	366	2	08BMQ4
13	58	51.3	398	2	06VS51
14	58	51.3	448	2	08XND2
15	57.5	50.9	699	1	SRCH_HUMAN
16	57	50.4	158	2	072942
17	57	50.4	294	2	07US12
18	57	50.4	315	2	08ATX7
19	57	50.4	322	2	08NR12
20	57	50.4	363	2	06E072
21	57	50.4	453	2	075C03
22	57	50.4	459	2	09PAU2
23	57	50.4	508	2	MARK_MARSC
24	57	50.4	508	2	09BBG0
25	57	50.4	508	2	09BRG3
26	57	50.4	630	2	08G074
27	57	50.4	1174	2	089NL8
28	56.5	50.4	1676	2	08TG36
29	56.5	50.0	105	2	06P8X5
30	56.5	50.0	211	2	06ZTC4
31	56.5	50.0	400	2	083A51

32	56	49.6	183	2	06D2A5	Q6d2a5 erwilia car
33	56	49.6	211	2	09N4A6	Q9n4a6 caenorhabdi
34	56	49.6	239	2	06LV35	Q6lv35 photobacter
35	56	49.6	242	2	058743	Q58743 echizosacch
36	56	49.6	270	2	09BLU4	Q9blu4 leishmania
37	56	49.6	296	2	08PHM1	Q8phm1 xanthomonas
38	56	49.6	313	2	09N4A7	Q9n4a7 caenorhabdi
39	56	49.6	331	2	067PX1	Q67px1 symbiodacte
40	56	49.6	348	2	06LXA4	Q6lxa4 methanococc
41	56	49.6	372	2	08MK09	Q8mk09 macaca mula
42	56	49.6	372	2	08BK10	Q8bk10 macaca mula
43	56	49.6	608	2	0910A3	Q910a3 cyprinus ca
44	56	49.6	2547	1	FAFX_HUMAN	Q93008 h probable
45	55.5	49.1	1749	2	08T0W6	Q8t0w6 echinococcu

ALIGNMENTS

```

RESULT 1
ID 062315 PRELIMINARY; PRT; 107 AA.
AC 062315;
DT 05-JUL-2004 (TREMblrel. 27, Created)
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
DE DCL protein-like.
GN Name=OJ1004_A11.16-2; Synonyms=P0539D10.35-2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_Taxid=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004817; BAD17127.1; -.
SQ
Sequence 107 AA; 12839 MW; 00F06AF3A12B0EB1 CRC64;

Query Match 60.2%; Score 68; DB 2; Length 107;
Best Local Similarity 37.5%; Pred. No. 1.3;
Matches 9; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

QY 3 HPOFEKXXXXXXXXXXSHPOFE 26
Db 30 HPOYEKKICGIDYLVGLHPEPE 53

RESULT 2
ID 08CJD2 PRELIMINARY; PRT; 704 AA.
AC 08CJD2;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Guanylyl cyclase alpha 1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura T., Suzuki N.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB096020; BAC24016.1; -.
DR HSSP; P30803; IAZS.
GO; GO:0004383; F:guanylate cyclase activity; IEA.

```

DR GO:0016829; F:lyase activity; IEA.
 DR GO:00007242; P:intracellular signalling cascade; IEA.
 DR InterPro; IPR001054; G_cyclase.
 DR Pfam; PF00211; Guanylate_cyc; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
 KW Lyase.
 SQ SEQUENCE 704 AA; 7918 MW; F73FC97B685446A CRC64;
 RT
 Query Match 59.3%; Score 67; DB 2; Length 704;
 Best Local Similarity 45.0%; Pred. No. 15;
 Matches 9; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Qy 8 KXXXXXXXXXXXXXMSHPQEF 27
 Db 685 KASGVLDVPRGWSHPQEF 704

RESULT 3

ID Q8CH90 PRELIMINARY; PRT; 743 AA.
 AC Q8CH90;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Nakamura I., Yao Y., Suzuki N.;
 RA Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB097860; BAC44887.1; -.
 DR HSSP; P30803; IAZS.
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR001054; G_cyclase.
 DR InterPro; IPR009080; CRNasey_1a_bind.
 DR Pfam; PF00211; Guanylate_cyc; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
 KW Lyase.
 SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match 54.9%; Score 62; DB 2; Length 743;
 Best Local Similarity 42.1%; Pred. No. 78;
 Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Qy 9 XXXXXXXXXXXXXXXXMSHPQEF 27
 Db 725 LKETSIVPRGWSHPQEF 743

RESULT 4

ID Q913C1 PRELIMINARY; PRT; 433 AA.
 AC Q913C1;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Probable cytochrome c.
 GN OrderedLocustNames=PA1600;
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OC NCBI_TaxID=287;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043; DOI=10.1038/35023079;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garner R.L., Goltzer L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Mu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RT Nature 406:958-964(2000).
 RU EMBL; AB04588; AAC04989.1; -.
 DR PIR; H63444; H63444.
 DR GO; GO:0004022; F:alcohol dehydrogenase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR003219; Cytc_adh.
 DR InterPro; IPR00345; Cytc_heme_BS.
 DR InterPro; IPR003056; Cytochrome_C.
 DR Pfam; PF00034; Cytochrom_C1.
 DR Prodom; PD01584; Cytc_adh_2.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_3.
 KW Complete proteome.
 SQ SEQUENCE 433 AA; 46248 MW; 83B8E0E75C5E6310 CRC64;

Query Match 54.0%; Score 61; DB 2; Length 433;
 Best Local Similarity 28.0%; Pred. No. 58;
 Matches 7; Conservative 15; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPQEFKXXXXXXXXXXXXXMSHPQEF 25
 Db 99 WSPYAFERAMRHGVARDGSYLVPF 123

RESULT 5

ID Q8RV8 PRELIMINARY; PRT; 224 AA.
 AC Q8RV8;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE DCU protein.
 GN Name=cl;
 OS Coffea arabica (Coffee).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Gentianales; Rubiaceae; Ixoroideae; Coffeae.
 OC NCBI_TaxID=13443;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bean;
 RA Marraccini P., Meunier A.;
 RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ420083; CAD12248.1; -.
 SQ SEQUENCE 224 AA; 25678 MW; E8B58F8A50393849E CRC64;

Query Match 53.1%; Score 60; DB 2; Length 224;
 Best Local Similarity 32.0%; Pred. No. 38;
 Matches 8; Conservative 13; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HPOFEKXXXXXXXXXXXXXMSHPQEF 27
 Db 151 HPOCEKKGSDVYITGYHDFDR 175

RESULT 6

ID Q9C642 PRELIMINARY; PRT; 219 AA.
 AC Q9C642;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```

DE 25-OCT-2004 (TREMBLrel. 28, last annotation update)
DE Defective chloroplasts and leaves (DCL) protein, putative
DE (A1g45261/F2G19.1) (At1g45261)
GN Name=F2G19.1;
OS Arabidopsis thaliana (Mouse-ear cress) .
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Matci R., Romling C.M., Koo H., Fujii C.Y., Uterback T.R.,
RA Barnesstead M.B., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Banh J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawaji T., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Ondexa C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Ecker J.R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC083835; AAC50632.1; -
DR EMBL; AY128277; AAAG1086.1; -
DR EMBL; BT014877; AAT41860.1; -
DR PIR; C96510; C96510.
SQ SEQUENCE 219 AA; 25448 MW; C4061B8DA2A4C448 CRC64;

Query Match 52.2%; Score 59; DB 2; Length 219;
Best Local Similarity 33.3%; Pred. No. 51;
Matches 8; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

Cy 3 HPOPEKXXXXXXXXXXMSHPQFE 26
Db 146 HPECEKKGCGIDYIMVGHHPDFE 169

RESULT 7
Q9AD61 PRELIMINARY; PRT; 389 AA.
ID Q9AD61
AC Q9AD61;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein SCP1.89c.
GN OrderedLocusNames=SCP1.89c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=AJ3(2) / M145;
RC MEDLINE=21996410; PubMed=1200953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kießer H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Croxall A., Fraser J., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kießer T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinovich E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor AJ3(2).";
RL Nature 417:141-147(2002).

```

Query Match	Best local Similarity	Score	DB 2;	Length	389;
Matches	8;	Conservative	15;	Mismatches	4;
			Indels		Gaps
DR EMBL; AL590463; CAC36610.1; -					
DR InterPro; IPR009061; Putativ_DNA_bind.					
KW Complete proteome; Hypothetical protein.					
SQ SEQUENCE 389 AA; 42800 MW; B27057143383AA27 CRC64;					
Query Match	52.2%;	Score 59;	DB 2;	Length 389;	
Best local Similarity	25.8%;	Pred. No. 97;			
Matches	8;	Conservative	15;	Mismatches	4;
			Indels		Gaps
Oy 1 WSHPOFEKXXXXXXXXXXW5----HPOFEK 27					
Db 266 WAHPEAVTFDGVSRGVGNGVAVHPQSD 236					
RESULT 8					
062TR3	PRELIMINARY;	PRT;	208 AA.		
AC Q62TR3.					
DT 25-OCT-2004 (TREMBLrel. 28, Created)					
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)					
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)					
DE Hypothetical protein.					
CN ORFNames=BL04042;					
OS Bacillus licheniformis DSM 13.					
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.					
OX NCBI_TaxID=279010;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=ATCC 14580;					
RA Rey M.W., Ramaiya P., Nelson B.A., Brody-Karpin S.D., Zaretsky E.J.,					
RA Tang M., de Leon A.U., Xiang H., Guertl V., Clausen I.G., Olsen P.B.,					
RA Rasmussen M.P., Andersen J.T., Jorgensen P.L., Larsen T.S.,					
RA Sorokin A., Bolotin A., Lapidus A., Galleron N., Ehrlich S.D.,					
RA Berka R.M.;					
RT "Complete genome sequence of the industrial bacterium Bacillus					
RT licheniformis and comparisons with closely related Bacillus species.";					
RL Genome Biol. 5:R77-R77(2004).					
DR EMBL; CP000002; AAU23846.1; -					
KW Hypothetical protein.					
SO SEQUENCE 208 AA; 23430 MW; 20E2053107BD86ED CRC64;					
Query Match	51.8%;	Score 58.5;	DB 2;	Length 208;	
Best local Similarity	22.2%;	Pred. No. 56;			
Matches	8;	Conservative	15;	Mismatches	4;
			Indels		Gaps
Oy 1 WSHPOFE-----KXXXXXXXXXXWSHPOFEK 27					
Db 153 WAHSEFTLDGALHEALRVARRVVKOHKWSPREK 188					
RESULT 9					
065IB2	PRELIMINARY;	PRT;	256 AA.		
AC Q65IB2.					
DT 25-OCT-2004 (TREMBLrel. 28, Created)					
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)					
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)					
DE Yp1P.					
CN Name=Yp1P; ORFNames=BL102322;					
OS Bacillus licheniformis DSM 13.					
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.					
OX NCBI_TaxID=279010;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=DSM 13;					
RC PubMed=15383718;					
RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H.,					
RA Ehrenreich P., Baumer S., Henne A., Liesegang H., Merkl R.,					
RA Ehrenreich A., Gottschalk G.;					
RT "The Complete Genome Sequence of Bacillus licheniformis DSM13, an					
RT Organism with Great Industrial Potential.";					
RT J. Mol. Microbiol. Biotechnol. 7:204-211(2004).					
RL EMBL; AE017333; AAU41202.1; -					

```
SO SEQUENCE      256 AA; 29029 MW; C127FF81C043FBA9 CRC64;

Query Match          51.8%; Score 58.5; DB 2; Length 256;
Best Local Similarity 22.2%; Pred. No. 71;
Matches              8; Conservative 15; Mismatches   4; Indels    9; Gaps     1

Qy                  1 WSHPOFE-----KXXXXXXXXXXMSHPQFK 27
                   ||| |
Db                    : ::::::::::::::||::|
                        201 WAHSEFTLDGALHEALRVARKVVLVDHWKSFRFX 236

RESULT 10
QHXS49
ID QHX549 PRELIMINARY; PRT; 103 AA.
AC OXHS49; O7ADV0;
DT 01-MAR-2002 (TREMBlrel.. 20, Created)
DT 01-MAR-2002 (TREMBlrel.. 20, Last sequence update)
DE Hypothetical protein Z2097 (Hypothetical protein ECs2201).
GN OrderedlocusNames=ECGs2201_22097;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacterales; Gammaproteobacteriales; Enterobacteriaceae;
OX NCBI_Taxid=83334;
PI [ ]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Pena N.T., Blunkett G. III, Burland V., Mau B., Glaser J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor U., Kirkpatrick H.A.,
RA Postal G., Hackert U., Link S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
RA Apodaca J.V., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7." ;
RL Nature 409:529-533(2001).[2]
RN RN
RM RM
RS RS
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952 / EHEC;
RX MEDLINE=2115631; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Oketani E., Nakayama K., Murata T., Tanaka M., Tobo T.,
RA Iida T., Takami H., Honda T., Saekawa C., Ogasaara N., Yasunaga T.,
RA Kuhara S., Shibata T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005346; AAC56167.1; -.
DR EMBL; AP002557; BAB35624.1; -.
DR PIR; A90904; A90904.
DR PIR; C85713; C85713.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11980 MW; 5A42AAF91CF29BE6 CRG64;

Query Match          51.3%; Score 58; DB 2; Length 103;
Best Local Similarity 30.0%; Pred. No. 30;
Matches             6; Conservative 14; Mismatches    0; Indels    0; Gaps     0;

Qy                8 KXXXXXXXKXXXXMSPHF 27
                   |:~::~||||
Db                 5 KEMPVARDGYGVWTHPEYEX 24

RESULT 11
QHXLVO
ID QHX4VO PRELIMINARY; PRT; 103 AA.
AC QHX4VO; O7ACU2;
DT 01-MAR-2002 (TREMBlrel.. 20, Created)
DT 01-MAR-2002 (TREMBlrel.. 20, Last sequence update)
DT 25-OCT-2004 (TREMBlrel.. 28, Last annotation update)
DN Hypothetical protein Z3120 (Hypothetical protein Ecgs2757).
GS OrderedlocusNames=ECS2757_Z3120;
OB Escherichia coli O157:H7.
```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927 / EHEC;
 RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054069;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller U.,
 RA Grodeck J.E., Davis N.W., Lim A., Dimalanta E.T., Potamovits K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
 RL Nature 409:529-533 (2001).
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / RIMD 0509952 / EHEC;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Okamoto K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-., Mabeu E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Saekawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RL "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AF005422; AAG57029.1; -.
 DR EMBL; AP002559; BAB36180.1; -.
 DR PIR; A85821; A85821.
 DR PIR; E90973; E90973.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 103 Aa; 12092 MW; D4F3CEA5A1089D30 CRC64;
 Query Match 51.3%; Score 58; DB 2; Length 103;
 Best Local Similarity 30.0%; Pred. NO. 30;
 Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;
 Db 8 KXXXXXXXXXXSHPOPEK 27
 5 KEMPERNEYGCMTHPEYK 24
 RESULT 12
 Q88MQ4 PRELIMINARY; PRT; 366 AA.
 AC Q88MQ4.
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE RND membrane fusion protein.
 GN OrderedLocustNames=P1516;
 OS Pseudomonas putida (strain KT2440).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=160488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2423060; PubMed=12534463;
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
 RA Martins dos Santos V.A.P., Fouts D.B., Gill S.R., Pop M., Holmes M.,
 RA Brinkac L.M., Beaman M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
 RA Haidu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
 RA Mancini J., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
 RA Moazzaz A., Ullrich T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
 RA Wedler H., Hauber J., Stepanich D., Hohnesiel J., Straetz M., Heim S.,
 RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuelmler B.,
 RA Fraser C.M.;
 RL "Complete genome sequence and comparative analysis of the
 RL metabolically versatile Pseudomonas putida KT2440.";
 RL Environ. Microbiol. 4:799-808 (2002).
 DR EMBL; AF016779; AAN67137.1; -.
 DR TIGR; P1516; -.
 GO; GO:0016020; C:membrane, IEA.

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M60052; AAA88071.1; -

DR PIR; A54660; A54660.

DR Genew; HGNC:5178; HRC.

DR MIM; 142705; -

DR GO; GO:0005509; F:calcium ion binding; TAS.

DR GO; GO:0006936; P:muscle contraction; TAS.

KW Calcium-binding; Polymorphism; Repeat; Signal.

FT SIGNAL 1 28

FT CHAIN 29 699

FT DOMAIN 193 204

FT DOMAIN 246 261

FT DOMAIN 106 365

FT REPEAT 180 213

FT REPEAT 238 270

FT REPEAT 295 318

FT REPEAT 343 365

FT DOMAIN 106 342

FT REPEAT 106 121

FT REPEAT 134 154

FT REPEAT 155 177

FT REPEAT 214 237

FT REPEAT 271 294

FT REPEAT 319 342

FT DOMAIN 627 673

FT VARIANT 96 96

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

FT VARIANT 204 204

SEQUENCE 699 AA; 80244 MM; 9922EEDF012C61DD CRC64;

Query Match 50.9%; Score 57.5; DB 1; Length 699;

Best Local Similarity 36.0%; Pred. No. 3e+02;

Matches 9; Conservative 11; Mismatches 4; Indels 1; Gaps 1;

QY 3 HPQFEKXXXXXXXXXXXXXGHPQFEK 27

DB 67 HPD-EKKDVSTENGHHFWSPDREK 90

Search completed: March 2, 2005, 12:44:30
 Job time : 101.317 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 121.61 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ15
Perfect score: 113
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfex 27

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq15dec04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	87.6	35	6	ABP60369
2	98.5	87.2	24	6	ABP60362
3	98.5	87.2	36	6	ABP60370
4	69.5	61.5	24	6	ABP60363
5	67.5	59.7	646	8	ADS20251
6	67.5	59.7	858	8	ADS20230
7	67	59.3	490	4	AAM40759
8	66	58.4	485	8	ADS20249
9	66	58.4	697	8	ADS20227
10	63.5	56.2	179	5	AAB24889
11	63	55.8	448	4	AAB46425
12	62.5	55.3	236	8	ADO19051
13	62.5	55.3	236	8	ADO19053
14	62	54.9	117	5	AAU97558
15	62	54.9	117	5	AAU97553
16	62	54.9	117	5	AAU97557
17	62	54.9	117	5	AAU97559
18	62	54.9	117	5	AAU97555
19	62	54.9	117	5	AAU97556
20	62	54.9	117	5	AAU97552
21	62	54.9	117	5	AAU97560
22	62	54.9	118	5	AAU97554
23	62	54.9	183	6	ADA27294
24	62	54.9	183	6	ADA00703
25	62	54.9	199	7	ADB87309

ALIGNMENTS

26	54.9	205	2	AAW93967	AAW93967 Plasmid p
27	54.9	209	6	ADA00709	Ada00709 Modified
28	54.9	248	7	ADB87310	ADB87310 Apolipop
29	54.9	254	2	AAW93969	AAW93969 Plasmid p
30	54.9	254	4	AAB46421	AAB46421 Bilin bin
31	54.9	258	6	ADA27293	Ada27293 Plasmid p
32	54.9	258	6	ADA00702	Ada00702 Modified
33	54.9	267	8	ADR70320	ADR70320 Polioviru
34	54.9	268	8	ADR70385	ADR70385 Polioviru
35	54.9	268	8	ADR70319	ADR70319 Polioviru
36	54.9	268	8	ADR70386	ADR70386 Polioviru
37	54.9	269	8	ADR28054	ADR28054 NPB polyP
38	54.9	269	8	ADS17525	ADS17525 Amino aci
39	54.9	273	8	ADO25157	ADO25157 Melanoma
40	54.9	274	8	ADO25153	ADO25153 Melanoma
41	54.9	275	8	ADO25154	ADO25154 Melanoma
42	54.9	275	8	ADO25155	ADO25155 Melanoma
43	54.9	276	8	ADO25149	ADO25149 Melanoma
44	54.9	277	8	ADO25150	ADO25150 Melanoma
45	54.9	277	8	ADO25152	ADO25152 Melanoma

RESULT 1	ABP60369	standard; peptide; 35 AA.
ID	ABP60369	
AC	ABP60369;	
DT	28-MAR-2003	(first entry)
DE	Streptavidin binding peptide SEQ ID NO 10.	
XX	Streptavidin; protein chip; microtitre plate; detection.	
OS	Synthetic.	
FH	Key	Location/Qualifiers
FT	Misc-difference 9..27	/label= unknown
FT		/note= "optionally deleted for 1-15 residues"
XX	DE1013776-AL.	
XX	02-OCT-2002.	
PF	21-MAR-2001; 2001DE-01013776.	
PR	21-MAR-2001; 2001DE-01013776.	
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX	Schmidt T;	
XX	WPI; 2003-031166/03.	
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
XX	Claim 7; Page 16; 18pp; German.	
PS	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it	

CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
XX
SQ Sequence 35 AA;

Query Match 87.6%; Score 99; DB 6; Length 35;
Best Local Similarity 77.1%; Pred. No. 8.8e-06;
Matches 27; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 27
DB 1 WSHPOFEKXXXXXXXXXXXXXXXXXXXXXWSHPOFEK 35

RESULT 2

ABP60362
ID ABP60362 standard; peptide; 24 AA.

AC ABP60362;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 3.

XX Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

PH Key Location/Qualifiers

FT MISC-difference 9.16
FT /label= unknown

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

DR WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.

PS Disclosure; Page 4; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
XX
SQ Sequence 24 AA;

Query Match 87.2%; Score 98.5; DB 6; Length 24;

Best Local Similarity 88.9%; Pred. No. 5.9e-06;
Matches 24; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
QY 1 WSHPOFEKXXXXXXXXXXXXXWSHPOFEK 27
DB 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 24

RESULT 3

ABP60370
ID ABP60370 standard; peptide; 36 AA.

AC ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

XX Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

PH Key Location/Qualifiers

FT Region /note= "GGGS repeats 2-5 optionally absent, residues 13-
FT 28"
FT 9.12
FT /label= GGGS_repeat

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

DR WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.

PS Claim 8; Page 16; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
XX
SQ Sequence 36 AA;

Query Match 87.2%; Score 98.5; DB 6; Length 36;
Best Local Similarity 44.4%; Pred. No. 1.1e-05;
Matches 16; Conservative 11; Mismatches 0; Indels 9; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 27
DB 1 WSHPOFEKGGGGGGGGGGGGGGGGGGGGGGGGWSHPOFEK 36


```
RESULT 4
ABP60363
ID ABP60363 standard; peptide; 24 AA.
XX
XX AC ABP60363;
XX
XX 28-MAR-2003 (first entry)
XX
XX Streptavidin binding peptide SEQ ID NO 4.
XX
XX Streptavidin; protein chip; microtitre plate; detection.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Misc-difference 9..18
XX /label= unknown
XX
XX Misc-difference 22..24
XX /label= unknown
XX
XX DE10113776-A1.
XX
XX 02-OCT-2002.
XX
XX 21-MAR-2001; 2001DE-01013776.
XX
XX 21-MAR-2001; 2001DE-01013776.
XX
XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX Schmidt T;
XX
XX WPI; 2003-031166/03.
XX
XX New isolated peptide, useful as affinity purification tag for recombinant
XX protein, comprises at least two high-affinity streptavidin-binding
XX modules.
XX
XX Disclosure; Page 4; 18pp; German.
XX
XX The invention relates to an isolated peptide (I) comprising at least two
XX individual modules separated by 0-50 amino acids, with each containing at
XX least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
XX streptavidin binding modules, are useful as affinity handles for
XX purification of recombinant fusion proteins (FP), also for detecting FP,
XX e.g. on protein chips or microtitre plates. The modules in (I) bind
XX strongly to streptavidin, with a co-operative effect that provides
XX stronger binding than a single tag, but are displaced by a competitor.
XX (I) does not interfere with the function of attached proteins (II) (so it
XX may not be essential to remove it); facilitates detection and has easily
XX controllable binding properties. (I) is particularly used for purifying
XX FP from dilute solution in batch formats (which use simpler apparatus
XX than column methods and result in lower loss of FP). The present sequence
XX is that of a streptavidin binding peptide disclosed with the invention
XX
XX Sequence 24 AA:
XX
XX Query Match 61.5%; Score 69.5; DB 6; Length 24;
XX Best Local Similarity 87.5%; Pred. No. 0.07;
XX Matches 21; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
```

```
XX
XX 18-NOV-2004 (first entry)
XX
XX Human aggrecanase modified ADAMTS4 (MTS4) protein - SEQ ID 49.
XX
XX ADAMTS4; a disintegrin-like and metalloprotease;
XX thrombospondin type 1 motif 4; repolyisin; zinc metalloprotease;
XX aggrecanase; osteopathic; antiinflammatory; antiarthritic; antineumatic;
XX cytosolic; osteoarthritis; glioma; cancer; inflammatory joint;
XX rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
XX human; chromosome 1q21-q23; enzyme.
XX
XX Homo sapiens.
XX
XX Synthetic.
XX
XX WO2004011637-A2.
XX
XX 05-FEB-2004.
XX
XX 29-JUL-2003; 2003WO-US023484.
XX
XX 29-JUL-2002; 2002US-0398721P.
XX
XX (AMHP) WYETH.
XX (CORC) CORCORAN C J.
XX (PLAN) PLANNERY C R.
XX (ZENG) ZENG W.
XX (RACI) RACIE L A.
XX (MCDO) MCDONAGH T.
XX (FREE) FREEMAN B A.
XX (GEOR) GEORGADIS K E.
XX (LAVA) LAVALLIE E R.
XX
XX Corcoran CJ, Plannery CR, Zeng W, Racie LA, McDonagh T;
XX Freeman BA, Georgiadis KE, Lavallie ER;
XX
XX WPI; 2004-143860/14.
XX
XX New isolated, modified ADAMTS4 (aggrecanase) protein with improved
XX stability useful for identifying inhibitors of the enzyme activity for
XX treating aggrecanase-associated conditions, including osteoarthritis.
XX
XX Claim 9; SEQ ID NO 49; 117pp; English.
XX
XX The invention relates to a novel isolated, modified ADAMTS4 (a
XX disintegrin-like and metalloprotease (repolysin type) with
XX thrombospondin type 1 motif 4) protein with improved stability compared
XX to a naturally occurring, full-length ADAMTS4 protein, where the modified
XX protein differs from the naturally-occurring, full-length ADAMTS4 protein
XX by at least one amino acid. ADAMTS proteins are a subfamily of zinc
XX metalloproteases and include aggrecanases amongst their members. The
XX protein of the invention demonstrates osteopathic, antiinflammatory,
XX antiarthritic, antineumatic and cytosolic activities and may be useful
XX for treating aggrecanase-associated conditions, including osteoarthritis,
XX glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
XX arthritis, periodontal disease and Crohn's disease. The current sequence
XX is that of a human aggrecanase modified ADAMTS4 (MTS4) protein of the
XX invention.
XX
XX Sequence 646 AA:
XX
XX Query Match 59.7%; Score 67.5; DB 8; Length 646;
XX Best Local Similarity 31.2%; Pred. No. 17;
XX Matches 10; Conservative 14; Mismatches 3; Indels 5; Gaps 1;
```

```
RESULT 5
ADS20251
ID ADS20251 standard; protein; 646 AA.
XX
XX ADS20251;
```

```
RESULT 6
ADS20230
ID ADS20230 standard; protein; 858 AA.
```

```

XX AC ADS20230;
XX XX
XX DT 18-NOV-2004 (first entry)
XX DE Human aggrecanase ADAMTS4 truncated protein with insert/Strep tag.
XX KW ADAMTS4, a disintegrin-like and metalloprotease;
XX KW thrombospondin type 1 motif 4; reprolysin; zinc metalloprotease;
XX KW cytosolic; osteopathic; antiinflammatory; zinc metalloprotease;
XX KW rheumatoid arthritis; glioma; cancer; inflammatory joint;
XX KW human; enzyme; chromosome 12q1-q23; truncation; Strep tag; mutant;
XX OS Homo sapiens.
XX OS Synthetic.
XX OS Unidentified.
XX FT Key
XX FT Location/Qualifiers
XX FT /note="Flexible spacer/enterokinase recognition site
XX FT /note="858
XX FT /note="Strep tag"
XX PN WO2004011337-A2.
XX PD 05-FEB-2004.
XX PF 29-JUL-2003; 2003WO-US023484.
XX PR 29-JUL-2002; 2002US-0398721P.
XX PA (AMBP) WYETH.
XX PA (CORC) CORCORAN C J.
XX PA (ZENG) FLANNERY C R.
XX PA (ZENG) ZENG W.
XX PA (RACE) RACE L A.
XX PA (MCD) MCDONAGH T.
XX PA (PRE) FREEMAN B A.
XX PA (GEOR) GEORGIADIS K E.
XX PA (LAVA) LAVALLIE E R.
XX PI Corcoran C J, Flannery CR, Zeng W, Race LA, McDonagh T,
XX PI Freeman BA, Georgiadis KE, Lavallie ER,
XX DR WPI; 2004-143860/14.
XX PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
XX PT stability useful for identifying inhibitors of the enzyme activity for
XX PT treating aggrecanase-associated conditions, including osteoarthritis.
XX PS Claim 9; SEQ ID NO 27; 117bp; English.
XX CC The invention relates to a novel isolated, modified ADAMTS4 (a
XX CC disintegrin-like and metalloprotease (reprolysin type) with
XX CC thrombospondin type 1 motif 4) protein with improved stability
XX CC to a naturally occurring, full-length ADAMTS4 protein with
XX CC by at least one amino acid ADAMTS4 protein, where the modified
XX CC metalloprotease and include ADAMTS4 proteins, a full-length ADAMTS4 protein
XX CC protein of the invention demonstrates osteopathic, antiinflammatory,
XX CC anarthritic, antineoplastic and cytostatic activities, their members. The
XX CC for treating aggrecanase-associated conditions, including osteoarthritis,
XX CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
XX CC arthritis, periodontal disease and Crohn's disease. The current sequence
XX CC is that of the human aggrecanase ADAMTS4 truncated protein with insert
XX CC peptide and Strep tag of the invention.
XX SQ Sequence 858 AA;
Query Match 59.7%; Score 67.5; DB 8; Length 858;

```

seq15.rag

```

Best Local Similarity 31.2%, Pred. No. 26;
Matches 10; Conservative 14; Mismatches 3; Indels 5; Gaps 1.
QY 1 MSHPQPE-----KXXXXXXXKXWMSHPQPEK 27
Db 827 WMRRAQQLTILRRRWRAGRKASMSHPQPEK 858
RESULT 7
AAM40759
ID AAM40759 standard; protein; 490 AA.
XX AC AAM40759;
XX AC
XX AC
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 5690.
XX KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukemia.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US034263.
XX PR 23-DEC-1999; 99US-00471275.
XX PR 21-JAN-2000; 2000US-00488725.
XX PR 20-APR-2000; 2000US-00582317.
XX PR 19-JUN-2000; 2000US-00598042.
XX PR 03-AUG-2000; 2000US-00620312.
XX PR 14-SEP-2000; 2000US-00653450.
XX PR 18-OCT-2000; 2000US-00662191.
XX PR 29-NOV-2000; 2000US-00693036.
XX PA (HYSB-) HYSBQ INC.
XX PI Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;
XX PI Zhou P, Goodrich R, Drmanac RT;
XX DR WPI; 2001-442253/47.
XX DR N-PSDB; AAI59915.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders such
XX PT as central nervous system injuries.
XX PS Example 2; SEQ ID NO 5690; 10078bp; English.
XX CC The invention relates to human nucleic acids (AAI5798-AAI1369) and the
XX CC encoded polypeptides (AAM38642-AAM42213) with neotropic.
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilization of the activities, chemokine/chemokine activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukemias and
XX CC part of the printed specification

```


PF 08-JUN-2000; 2000WO-DE001873.
 XX
 PR 08-JUN-1999; 99DE-01026068.
 XX
 PA (SKER/) SKERRA A.
 XX
 PI Skerra A, Schlehuber S;
 XX
 DR WPI; 2001-071071/08.
 DR N-PSDB; AAF25710.
 XX
 PT New muteins of bilin-binding protein, useful for detecting digoxigenin
 PT being used as label in e.g. binding assays, are very selective for
 PT digoxigenin.
 XX
 PS Example 3; Page 65-67; 80pp; German.
 XX
 CC This invention describes novel polypeptides (I) that are muteins of bilin
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
 CC have an amino acid (aa) substitution at at least one of the positions 28,
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
 CC fusion proteins, are used to bind, detect, determine, immobilize or
 CC separate Dig or its conjugates with proteins, nucleic acids,
 CC carbohydrates, other biological or synthetic macromolecules or low
 CC molecular weight compounds, particularly in assays where Dig is being
 CC used as a label. Compared with Dig-specific antibodies, (I) have a
 CC simpler structure and are easier to prepare. They have very high
 CC specificity for Dig, relative to other steroids, and fusion partners may
 CC be attached to either end without compromising their ability to bind
 CC ligand
 XX
 SQ Sequence 448 AA;
 XX
 Query Match 55.8%; Score 63; DB 4; Length 448;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 10; Conservative 10; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSHPOFEKXXXXXXXXXXW 20
 DB 198 WSHPOFEKXXASGRFMKGF 217
 XX
 RESULT 12
 ADO19051
 ID ADO19051 standard; protein; 236 AA.
 XX
 AC ADO19051;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Murine antibody M1 Fab fragment heavy chain.
 XX
 KW Mouse; antibody; M1 Fab; heavy chain; Fab; scFv; abused drug; morphine;
 KW THC; amphetamine; environmental hazard; toxic compound;
 KW microbial process; metabolic process; drug monitoring;
 KW pharmacological research.
 XX
 OS Mus sp.
 XX
 PN WO2004046733-A1.
 XX
 PD 03-JUN-2004.
 XX
 PF 17-NOV-2003; 2003WO-FI000875.
 XX
 PR 18-NOV-2002; 2002FI-00002048.
 XX
 PA (VALM) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
 XX
 PI Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H;
 XX
 PS WPI; 2004-420710/39.
 XX

XX
 PT Non-competitive immunoassay for small analyte, useful for assaying drug
 PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
 PT reagent pair comprising a first binding partner and a second binding
 PT partner.
 XX
 PS Claim 21; SEQ ID NO 2; 35pp; English.
 XX
 CC The invention relates to a non-competitive immunoassay for a small
 CC analyte, comprising reacting a sample containing the analyte with a
 CC reagent pair comprising a first binding partner that binds to the analyte
 CC and a second binding partner that binds to the complex of the analyte and
 CC the first binding partner, and determining the binding of the second
 CC binding partner, thus indicating the presence of the analyte in the
 CC sample. The first and second binding partners are antibody fragments Fab
 CC or scFv. The reagent pair is useful in a non-competitive immunoassay for
 CC a small analyte, particularly for assaying drugs of abuse e.g., morphine,
 CC THC or amphetamine. The immunoassay is useful for detecting environmental
 CC hazards, toxic compounds in food and feed, chemicals indicative of
 CC ongoing processes (e.g., microbial processes in buildings, metabolic
 CC processes of living organisms) and in clinical tests, drug monitoring and
 CC pharmacological research. This sequence represents the murine antibody M1
 CC Fab fragment heavy chain, used in the method of the invention.
 XX
 SQ Sequence 236 AA;
 XX
 Query Match 55.3%; Score 62.5; DB 8; Length 236;
 Best Local Similarity 34.5%; Pred. No. 19;
 Matches 10; Conservative 12; Mismatches 4; Indels 3; Gaps 1;
 QY 2 SHPOFE---KXXXXXXXXXXWSHPOFEK 27
 DB 208 AHPASTKVDKXIVRDCGTSMSHPOFEK 236
 XX
 RESULT 13
 ADO19053
 ID ADO19053 standard; protein; 236 AA.
 XX
 AC ADO19053;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Murine antibody M2 Fab fragment heavy chain.
 XX
 KW Mouse; antibody; M2 Fab; heavy chain; Fab; scFv; abused drug; morphine;
 KW THC; amphetamine; environmental hazard; toxic compound;
 KW microbial process; metabolic process; drug monitoring;
 KW pharmacological research.
 XX
 OS Mus sp.
 XX
 PN WO2004046733-A1.
 XX
 PD 03-JUN-2004.
 XX
 PF 17-NOV-2003; 2003WO-FI000875.
 XX
 PR 18-NOV-2002; 2002FI-00002048.
 XX
 PA (VALM) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
 XX
 PI Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H;
 XX
 PS WPI; 2004-420710/39.
 XX
 PT Non-competitive immunoassay for small analyte, useful for assaying drug
 PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
 PT reagent pair comprising a first binding partner and a second binding
 PT partner.
 XX
 PS Claim 21; SEQ ID NO 4; 35pp; English.
 XX

CC The invention relates to a non-competitive immunoassay for a small
CC analyte, comprising reacting a sample containing the analyte with a
CC reagent pair comprising a first binding partner that binds to the analyte
CC and a second binding partner that binds to the complex of the analyte and
CC the first binding partner, and determining the binding of the second
CC binding partner, thus indicating the presence of the analyte in the
CC sample. The first and second binding partners are antibody fragments Fab
CC or scFv. The reagent pair is useful in a non-competitive immunoassay for
CC a small analyte, particularly for assaying drugs of abuse e.g., morphine,
CC THC or amphetamine. The immunoassay is useful for detecting environmental
CC hazards, toxic compounds in food and feed, chemicals indicative of
CC ongoing processes (e.g., microbial processes in buildings, metabolic
CC processes of living organisms) and in clinical tests, drug monitoring and
CC pharmacological research. This sequence represents the murine antibody M2
CC Fab fragment heavy chain, used in the method of the invention.

XX Sequence 236 AA;

Query Match 55.3%; Score 62.5; DB 8; Length 236;
Best Local Similarity 34.5%; Pred. No. 19;
Matches 10; Conservative 12; Mismatches 4; Indels 3; Gaps 1;

QY 2 SHPOFE--KXXXXXXXSHPOFEK 27
:|||||:|||||
Db 208 AHPASSTRKVKIIVPRDGTSMHPQFEK 236

RESULT 14

AAU97558
ID AAU97558 standard; protein; 117 AA.

AC AAU97558;

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #6.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;

KW irrigation water; waste stream; contaminated aqueous medium;

KM biological fluid; gastrointestinal tract; chelon protein;

KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;

XX heavy metal binding protein.

OS Synthetic.

PN WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819.

PR 12-OCT-2000; 2000US-0240465P.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Summers AO, Caguiat JI;

PI WPI; 2002-435437/46.

DR Novel non-naturally occurring recombinant DNA molecule encoding a chelon

XX protein useful for binding divalent cation mercury from contaminated

XX soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

XX The present invention relates to a new non-naturally occurring

XX recombinant DNA molecule comprising a sequence encoding a chelon protein

XX which binds mercuric ions. The invention is useful for recombinantly

XX producing a protein in a host-cell, by infecting or transforming a host

XX cell capable of expressing a chelon coding sequence with a vector

XX comprising a promoter active in the host cell operably linked to a coding

XX region for the protein to produce a recombinant host cell and culturing

XX the recombinant host cell under conditions, where DNA is expressed. The

CC nucleic acid encoding the chelon protein is useful for binding divalent
CC mercuric ions, to take up, sequester and concentrate the heavy metal ions
CC from contaminated soil, ground water, hydroponic solutions or irrigation
CC water of waste streams. The DNA of the invention, when immobilised onto a
CC solid support, is useful for concentrating heavy metal ions from
CC contaminated environment waste streams or contaminated aqueous medium
CC including biological fluids. The nucleic acid, when recombinantly
CC expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic),
CC is suitable for use in the in vivo sequestration and elimination of
CC mercuric ion from gastrointestinal tracts of animals or humans exposed to
CC toxic metal ions such as mercury and/or cadmium. The molecules of the
CC invention are also useful in water treatment resins. The nucleic acid of
CC the invention is highly specific and binds divalent cation such as
CC mercury or cadmium with high affinity. The present amino acid sequence
CC represents one of a collection (AAU97553-AAU97560) of synthetic
CC cadmium/mercury ion binding chelon proteins of the invention. This
CC sequence is one of the heavy metal binding proteins termed chelons of the
CC invention

XX Sequence 117 AA;

Query Match 54.9%; Score 62; DB 5; Length 117;
Best Local Similarity 42.1%; Pred. No. 8.2;
Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXSHPOFEK 27
:|||||:|||||
Db 99 ARKGNVSCPSAMSHPOFEK 117

RESULT 15

AAU97553
ID AAU97553 standard; protein; 117 AA.

AC AAU97553;

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #1.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;

KW irrigation water; waste stream; contaminated aqueous medium;

KM biological fluid; gastrointestinal tract; chelon protein;

KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;

XX heavy metal binding protein.

OS Synthetic.

PN WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819.

PR 12-OCT-2000; 2000US-0240465P.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Summers AO, Caguiat JI;

PI WPI; 2002-435437/46.

DR Novel non-naturally occurring recombinant DNA molecule encoding a chelon

XX protein useful for binding divalent cation mercury from contaminated

XX soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

XX The present invention relates to a new non-naturally occurring

XX recombinant DNA molecule comprising a sequence encoding a chelon protein

XX which binds mercuric ions. The invention is useful for recombinantly

XX producing a protein in a host-cell, by infecting or transforming a host

XX cell capable of expressing a chelon coding sequence with a vector

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 83.9634 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ15
Perfect score: 113
Sequence: 1 wshpgefxxxxxxwshpgefek 27

Scoring table: BLOSUM62GX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing filer 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
18:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98.5	87.2	24	US-10-026-578B-3	Sequence 3, Appl1
2	98.5	87.2	36	US-10-026-578B-10	Sequence 10, Appl1
3	85.5	75.7	36	US-10-026-578B-11	Sequence 11, Appl1
4	69.5	61.5	24	US-09-809-517A-31	Sequence 31, Appl1
5	69.5	61.5	24	US-10-026-578B-4	Sequence 4, Appl1
6	69.5	61.5	25	US-09-809-517A-34	Sequence 34, Appl1
7	69.5	61.5	245	US-10-887-228A-1	Sequence 1, Appl1
8	69.5	61.5	246	US-10-887-228A-9	Sequence 9, Appl1
9	69.5	61.5	252	US-10-887-228A-5	Sequence 5, Appl1
10	68	60.2	134	US-10-437-963-104065	Sequence 104065,
11	68	60.2	158	US-10-424-599-184769	Sequence 184769,
12	68	60.2	212	US-10-424-599-184075	Sequence 184075,
13	68	60.2	252	US-10-437-963-200150	Sequence 200150,

14	67.5	59.7	646	US-10-628-432-49	Sequence 49, Appl1
15	67.5	59.7	858	US-10-628-432-27	Sequence 27, Appl1
16	66	58.4	485	US-10-628-432-47	Sequence 47, Appl1
17	66	58.4	697	US-10-628-432-24	Sequence 24, Appl1
18	64	56.6	21	US-09-809-517A-30	Sequence 30, Appl1
19	64	56.6	22	US-09-809-517A-33	Sequence 33, Appl1
20	62	54.9	117	US-09-977-137A-4	Sequence 4, Appl1
21	62	54.9	117	US-09-977-137A-5	Sequence 5, Appl1
22	62	54.9	117	US-09-977-137A-7	Sequence 7, Appl1
23	62	54.9	117	US-09-977-137A-8	Sequence 8, Appl1
24	62	54.9	117	US-09-977-137A-9	Sequence 9, Appl1
25	62	54.9	117	US-09-977-137A-10	Sequence 10, Appl1
26	62	54.9	117	US-09-977-137A-11	Sequence 11, Appl1
27	62	54.9	117	US-09-977-137A-12	Sequence 12, Appl1
28	62	54.9	118	US-09-977-137A-6	Sequence 6, Appl1
29	62	54.9	633	US-10-628-432-53	Sequence 53, Appl1
30	62	54.9	661	US-10-358-983-8	Sequence 8, Appl1
31	62	54.9	763	US-10-358-283-15	Sequence 15, Appl1
32	62	54.9	845	US-10-628-432-40	Sequence 40, Appl1
33	59.5	52.7	56	US-09-833-245-1103	Sequence 1103, Ap
34	59.5	52.7	56	US-09-833-245-1105	Sequence 1105, Ap
35	59	52.2	108	US-10-425-114-37298	Sequence 37298, A
36	58	51.3	792	US-10-425-963-180912	Sequence 180912,
37	57.5	50.9	699	US-10-408-765A-434	Sequence 434, App
38	57.5	50.9	699	US-10-741-600-851	Sequence 851, App
39	57.5	50.9	699	US-10-741-600-852	Sequence 852, App
40	57	50.4	36	US-10-424-599-146965	Sequence 146965,
41	57	50.4	385	US-10-437-963-115643	Sequence 115643,
42	57	50.4	472	US-09-369-735B-8	Sequence 8, Appl1
43	56	49.6	100	US-10-424-599-258116	Sequence 258116,
44	56	49.6	529	US-10-437-963-132697	Sequence 132697,
45	56	49.6	706	US-10-437-963-132699	Sequence 132699,

ALIGNMENTS

RESULT 1
US-10-026-578B-3
Sequence 3, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (10)..(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:

```

NAME/KEY: MISC_FEATURE
LOCATION: (12)-(12)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (13)-(13)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (14)-(14)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (15)-(15)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (16)-(16)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
IS-10-026-578b-3

```

Query Match	88.2%	Score 98.5	DB 14	Length 24
Best Local Similarity	88.9%	Pred. No. 4e-05		
Matches 24	Conservative	0	Mismatches	3
			Indels	1
			Gaps	

QY 1 WSHPQFEKXXXXXXXXXXXXXWSHPQFEK 27
Db 1 WSHPQFEK---XXXXXXXXXXWSHPQFEK 24

RESULT 2
US-10-026-578B-10
; Sequence 10, Application US/10026578B
; Publication No. US20030083474A1
; Chemical Information:

```

APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tags
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026.578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC FEATURE
LOCATION: (9)-(28)
OTHER INFORMATION: x represents a single amino acid at each of the positions indicated
OTHER INFORMATION: ed, some of the amino acids may be missing. Where amino acids are
OTHER INFORMATION: missing, the total numbers of x will be no less than 5
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10

```

Query Match	87.2%	Score 98.5;	DB 14;	Length 36;
Best Local Similarity	75.0%;	Pred. No. 6.8e-05;		
Matches 27; Conservative	0;	Mismatches 0;	Indels 9;	Gaps 1;

```

0Y      1 WSHPQFEK-----XXXXXXXXXXXXWSHPQFEK 27
        |||||          |||||

```

Db 1 WSHPQFEKXXXXXXXXXXXXXXXXXXXXWSHPQFEK 36

RESULT 3
US-10-026-578B-11
; Sequence 11, Application US/10026578B
; Publication No. US20030083474A1
GENERAL INFORMATION:

? APPLICANT: Schmidt, Thomas
 ? TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tra
 ? FILE REFERENCE: 100810.01US1
 ? CURRENT APPLICATION NUMBER: US/10/026,578B
 ? CURRENT FILING DATE: 2002-11-11
 ? PRIOR APPLICATION NUMBER: DE 101 13 776.1
 ? PRIOR FILING DATE: 2001-03-21
 ? PRIOR APPLICATION NUMBER: BCT/EP01/11846
 ? PRIOR FILING DATE: 2001-10-12
 ? NUMBER OF SEQ ID NOS: 14
 ? SOFTWARE: PatentIn version 3.1

```

; SEQ ID NO 11
; LENGTH: 36

```

```

; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Synthetic Peptide
;
; FEATURE:
;

```

NAME/KEY: MISC_FEATURE

OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more (GGGS) repeats may be missing. However, at least one (GGGS) repeat will be present

OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11

Query Match	75.7%;	Score 85.5;	DB 14;	Length 36;
Best Local Similarity	41.7%;	Pred. No. 0.0032;		
Matches 15;	Conservative 11;	Mismatches 1;	Indels 9;	Gaps 1;

OY 1 WSPDQFEK-----XXXXXXXXXXXXWSPDQFEK 27
 |||||
 ::::::::::: |||||

Db 1 WSPHQFEKGSGSGSGSGSGSGSQSHPDQFEK 36

RESULT 4
US-09-809-517A-31

```

; Sequence 31, Application U
; Patent No. US20020034733A1

```

```

/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. US2002003473A1el methods for displaying (poly)peptides/prote
/ TITLE OF INVENTION: particles via disulfide bonds
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ CURRENT FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ PRIOR FILING DATE: 2000-02-18
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 31
/ LENGTH: 24
/ rddr. dbp

```

```

; ORGANISM: artificial sequence
;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

```

Query Match	61.5%;	Score 69.5;	DB 9;	Length 24;
Best Local Similarity	45.8%;	Pred. No. 0.22;		

```

Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

QY 5 QFE-KXXXXXXXXXXMSHPQFEK 27
   :|||:|||||:|||||:|||||:
   :|||:|||||:|||||:|||||:
DB 1 EFEOKLISEEDLNGAPWMSHPQFEK 24

RESULT 5
US-10-026-578B-4
; Sequence 4, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; FEATURE:
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; NAME/KEY: MISC FEATURE
; LOCATION: (10)..(10)
; FEATURE:
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; NAME/KEY: MISC FEATURE
; LOCATION: (11)..(11)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (12)..(12)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (13)..(13)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (14)..(14)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (16)..(16)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (17)..(17)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; NAME/KEY: MISC FEATURE
; LOCATION: (22)..(22)

; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (23)..(23)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (24)..(24)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-4

Query Match 61.5%; Score 69.5; DB 14; Length 24;
Best Local Similarity 87.5%; Pred. No. 0.22;
Matches 21; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXMSHPQ 24
   :|||:|||||:|||||:|||||:
   :|||:|||||:|||||:|||||:
DB 1 WSHPOFEKXXXXXXXXXX--HPQ 21

RESULT 6
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match 61.5%; Score 69.5; DB 9; Length 25;
Best Local Similarity 45.8%; Pred. No. 0.23;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

QY 5 QFE-KXXXXXXXXXXMSHPQFEK 27
   :|||:|||||:|||||:|||||:
   :|||:|||||:|||||:|||||:
DB 2 EFEOKLISEEDLNGAPWMSHPQFEK 25

RESULT 7
US-10-887-228A-1
; Sequence 1, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S30569US
; CURRENT APPLICATION NUMBER: US/10/887,228A
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12

```

```
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 245
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-1

Query Match      61.5%; Score 69.5; DB 17; Length 245;
Best Local Similarity 45.8%; Pred. No. 4.4;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

Cy 5 QFE-KXXXXXXSHPOFEK 27
Db 222 EFEQKLISEDLNGAPWSPHPOFEK 245

RESULT 8
US-10-887-228A-9
/ Sequence 9, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ FILE REFERENCE: S30569US
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 9
/ LENGTH: 246
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match      61.5%; Score 69.5; DB 17; Length 246;
Best Local Similarity 45.8%; Pred. No. 4.4;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

Cy 5 QFE-KXXXXXXSHPOFEK 27
Db 223 EFEQKLISEDLNGAPWSPHPOFEK 246

RESULT 9
US-10-887-228A-5
/ Sequence 5, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ FILE REFERENCE: S30569US
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5
/ LENGTH: 252
/ TYPE: PRT
/ ORGANISM: artificial
```

```
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-5

Query Match      61.5%; Score 69.5; DB 17; Length 252;
Best Local Similarity 45.8%; Pred. No. 4.5;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

Cy 5 QFE-KXXXXXXSHPOFEK 27
Db 222 EFEQKLISEDLNGAPWSPHPOFEK 252

RESULT 10
US-10-437-963-104065
/ Sequence 104065, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 104065
/ LENGTH: 134
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_101434C.1.pap
US-10-437-963-104065

Query Match      60.2%; Score 68; DB 16; Length 134;
Best Local Similarity 37.5%; Pred. No. 3.1;
Matches 9; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

Cy 3 HPOFEKXXXXXXSHPOFE 26
Db 57 HPOYEKKIGGIDYLTGALPEFE 80

RESULT 11
US-10-424-599-184769
/ Sequence 184769, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J
/ APPLICANT: Kovalic, David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 184769
/ LENGTH: 158
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_137862C.1.pap
US-10-424-599-184769

Query Match      60.2%; Score 68; DB 15; Length 158;
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 30.7317 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ15

Perfect score: 113
Sequence: 1 whnpqfexkxxxxxxxxwhnpqfex 27

Scoring table: BLOSUM62GX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.5	61.5	24	US-09-809-517A-31	Sequence 31, Appl
2	69.5	61.5	25	US-09-809-517A-34	Sequence 34, Appl
3	64	56.6	21	US-09-809-517A-30	Sequence 30, Appl
4	64	56.6	22	US-09-809-517A-33	Sequence 33, Appl
5	62	54.9	117	US-09-977-137A-4	Sequence 4, Appl
6	62	54.9	117	US-09-977-137A-5	Sequence 5, Appl
7	62	54.9	117	US-09-977-137A-7	Sequence 7, Appl
8	62	54.9	117	US-09-977-137A-8	Sequence 8, Appl
9	62	54.9	117	US-09-977-137A-9	Sequence 9, Appl
10	62	54.9	117	US-09-977-137A-10	Sequence 10, Appl
11	62	54.9	117	US-09-977-137A-11	Sequence 11, Appl
12	62	54.9	117	US-09-977-137A-12	Sequence 12, Appl
13	62	54.9	118	US-09-977-137A-6	Sequence 6, Appl
14	61	54.0	434	US-09-252-991A-25931	Sequence 25931, A
15	60	53.1	386	US-08-895-707-7	Sequence 7, Appl
16	57.5	50.9	699	US-09-538-092-995	Sequence 995, App
17	56	49.6	125	US-09-621-976-4303	Sequence 4303, Ap
18	56	49.6	306	US-09-248-796A-16127	Sequence 16127, A
19	56	49.6	2547	US-09-058-489-35	Sequence 35, Appl
20	56	49.6	2547	US-09-538-092-1374	Sequence 1374, Ap
21	55.5	49.1	647	US-09-248-796A-15400	Sequence 15400, A
22	55	48.7	84	US-09-270-767-36127	Sequence 36127, A
23	55	48.7	84	US-09-270-767-51344	Sequence 51344, A
24	54	47.8	223	US-09-489-039A-12356	Sequence 12356, A
25	54	47.8	394	US-09-248-796A-17857	Sequence 17857, A
26	53	46.9	10	US-09-809-517A-6	Sequence 6, Appl
27	53	46.9	400	US-07-989-991A-2	Sequence 2, Appl

28	53	46.9	400	3	US-08-621-255-2	Sequence 2, Appl
29	53	46.9	400	3	US-09-352-574-2	Sequence 2, Appl
30	53	46.9	638	4	US-09-252-991A-25205	Sequence 25205, A
31	53	46.9	979	1	US-08-308-881-6	Sequence 6, Appl
32	53	46.9	979	2	US-09-058-263-6	Sequence 6, Appl
33	53	46.9	979	2	US-09-059-099-6	Sequence 6, Appl
34	53	46.9	979	3	US-09-058-264-6	Sequence 6, Appl
35	53	46.9	979	4	US-09-455-962-6	Sequence 6, Appl
36	53	46.9	979	5	PCT-US95-06530-6	Sequence 6, Appl
37	52.5	46.5	389	2	US-08-811-949-65	Sequence 65, Appl
38	52	46.0	27	2	US-08-310-912A-47	Sequence 47, Appl
39	52	46.0	27	3	US-08-841-089-47	Sequence 47, Appl
40	52	46.0	27	3	US-09-301-085-47	Sequence 47, Appl
41	52	46.0	27	5	PCT-US95-04570-47	Sequence 47, Appl
42	52	46.0	27	5	PCT-US95-04589-47	Sequence 47, Appl
43	52	46.0	249	4	US-09-270-767-48926	Sequence 48926, A
44	52	46.0	250	4	US-09-270-767-33709	Sequence 33709, A
45	52	46.0	659	4	US-09-252-991A-26013	Sequence 26013, A

ALIGNMENTS

```
RESULT 1
US-09-809-517A-31
Sequence 31, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: Patentin version 3.0
SEQ ID NO 31
LENGTH: 24
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31
Query Match      61.5%; Score 69.5; DB 4; Length 24;
Best Local Similarity 45.8%; Pred. No. 0.0088;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;
OY 5 OPE-KXXXXXXWHPQFEX 27
:|||||:
Db 1 EPEQKLSBEDNGAPWHPQFEX 24
RESULT 2
US-09-809-517A-34
Sequence 34, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: Patentin version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match
Best Local Similarity 61.5%; Score 69.5; DB 4; Length 25;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

QY 5 EFKXXXXXXXXXMSHPQPEK 27
Db 2 EFKDKLSEEDLNGAPMSHPQPEK 25

RESULT 3
US-09-809-517A-30.
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match
Best Local Similarity 56.6%; Score 64; DB 4; Length 21;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;

QY 7 EKKXXXXXXXXXMSHPQPEK 27
Db 1 EFDYKDDDDKGA PMSHPQPEK 21

RESULT 4
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match
Best Local Similarity 56.6%; Score 64; DB 4; Length 22;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;

QY 7 EKKXXXXXXXXXMSHPQPEK 27
Db 2 EFDYKDDDDKGA PMSHPQPEK 22

RESULT 5
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match
Best Local Similarity 54.9%; Score 62; DB 4; Length 117;
Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXMSHPQPEK 27
Db 99 ARKGNVSCPSAMSHPOPEK 117

RESULT 6
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match
Best Local Similarity 54.9%; Score 62; DB 4; Length 117;
Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;
```



```

; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

```

```

Query Match          54.9%; Score 62; DB 4; Length 117;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches      8; Conservative 11; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 27
       :XXXXXXXXXXXXXXXXXXXX:
Db      99 ARKGNVPCPSAMSHPOFEK 117

```

```

RESULT 12
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

```

```

Query Match          54.9%; Score 62; DB 4; Length 117;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches      8; Conservative 11; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 27
       :XXXXXXXXXXXXXXXXXXXX:
Db      99 ARKGNVPCPSAMSHPOFEK 117

```

```

RESULT 13
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18

```

```

; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

```

```

Query Match          54.9%; Score 62; DB 4; Length 118;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches      8; Conservative 11; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXSHPOFEK 27
       :XXXXXXXXXXXXXXXXXXXX:
Db      100 ARKGNVPCPSAMSHPOFEK 118

```

```

RESULT 14
US-09-252-991A-25931
; Sequence 25931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25931
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25931

```

```

Query Match          54.0%; Score 61; DB 4; Length 434;
Best Local Similarity 28.0%; Pred. No. 12;
Matches      7; Conservative 15; Mismatches      3; Indels      0; Gaps      0;

```

```

QY      1 WSHPOFEKXXXXXXXXXSHPOF 25
       :XXXXXXXXXXXXXXXXXXXX:
Db      100 WSYPAFERAMRHVARDGYLPAPF 124

```

```

RESULT 15
US-08-895-707-7
; Sequence 7, Application US/08895707
; Patent No. 6077700
; GENERAL INFORMATION:
; APPLICANT: (Pharmacia & Upjohn, Co.)
; APPLICANT: alternatively, for U.S. filing:
; APPLICANT: Hollingsworth, Robert A.
; APPLICANT: Sharma, Satish K.
; APPLICANT: Rank, Kenneth B.
; APPLICANT: Evans, David B.
; TITLE OF INVENTION: Special Constructs and Complexes of
; TITLE OF INVENTION: Cyclin E
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: MI
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Rel. #1.0, Ver. #1.25/WordPerfect 5.2+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,707
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wootten, Thomas A.
REGISTRATION NUMBER: 35,004
REFERENCE/DOCKET NUMBER: 6054
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-7914
TELEFAX: 616-833-8897
INFORMATION FOR SEQ. ID NO. 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-895-707-7

Query Match 53.1%; Score 60; DB 3; Length 386;
Best Local Similarity 136.4%; Pred. No. 14;
Matches 8; Conservative 11; Mismatches 3; Indels 0; Gaps 0;
Qy 4 POFKXXXXXXXXXXWSHPOF 25
Db 363 POSGKKQSSGPEMASAKRHPOF 384

Search completed: March 2, 2005, 12:25:39
Job time : 31.7317 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 21.3984 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ16
Perfect score: 114
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfex 28

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 79:.*
2: PIR1:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	52.6	435	2 A44308	Antho-RPamide prec
2	60	52.6	1107	2 T25450	hypothetical prote
3	59.5	52.2	601	2 T11677	probable transcrip
4	58	50.9	313	2 T40926	conserved hypothet
5	58	50.9	652	2 T34497	hypothetical prote
6	57	50.0	151	2 T00816	hypothetical prote
7	57	50.0	206	2 S76147	hypothetical prote
8	56.5	49.6	342	2 A45462	C 4.2.3.9 aristo
9	56	49.1	103	2 B90973	hypothetical prote
10	56	49.1	103	2 A85821	unknown protein en
11	56	49.1	103	2 C85713	unknown protein en
12	56	49.1	103	2 A90904	hypothetical prote
13	56	49.1	334	2 A39172	Antho-RPamide neur
14	56	49.1	335	2 AD3492	acetoacetate-CoA 1
15	56	49.1	583	2 S30014	hypothetical prote
16	56	49.1	699	2 A54660	histidine rich cal
17	56	49.1	732	2 H83376	1,4-alpha-glucan b
18	56	49.1	2408	2 T24483	hypothetical prote
19	55.5	48.7	213	2 AD2533	hypothetical prote
20	55	48.2	234	2 T30473	late expression fa
21	55	48.2	253	2 AG2559	transposase al1807
22	55	48.2	398	2 AB2622	succinyl-diaminop
23	55	48.2	398	2 B97404	succinyl-diaminop
24	55	48.2	527	2 C88042	protein P56D12.4 l
25	55	48.2	624	2 T26148	hypothetical prote
26	55	48.2	995	2 T39724	probable pre-mRNA
27	55	48.2	3071	2 T45584	hypothetical prote
28	54	47.4	317	2 S49005	non-structural pro
29	54	47.4	323	2 B70960	probable nadr prot

30	54	47.4	484	1 SYRZET	glutamate-tRNA lig
31	54	47.4	507	2 T24944	hypothetical prote
32	54	47.4	537	2 YRMSB6	tyrosinase-related
33	54	47.4	617	2 F75484	hypothetical prote
34	54	47.4	685	2 E82297	c-di-GMP phosphodi
35	54	47.4	801	2 H83737	glucosidase BH0704
36	53.5	46.9	211	2 B71058	hypothetical prote
37	53.5	46.9	553	1 S29861	hybrid cluster 14P
38	53.5	46.9	1091	2 S01998	contractin precuro
39	53	46.5	139	2 T41693	conserved hypothet
40	53	46.5	228	2 S73041	hypothetical prote
41	53	46.5	230	2 B83879	hypothetical prote
42	53	46.5	246	2 C83694	hypothetical prote
43	53	46.5	282	2 T27554	hypothetical prote
44	53	46.5	290	2 T43351	nuclear receptor N
45	53	46.5	297	2 A45442	transport versicle

ALIGNMENTS

RESULT 1
A44308
Antho-RPamide precursor - sea anemone (Anthopleura elegantissima)
C:Species: Anthopleura elegantissima
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Jul-2004
C:Accession: A44308
R:Schmutterler, C.; Darner, D.; Diekhoff, D.; Grimmelikhuijzen, C.J.
A:Title: Identification of a novel type of processing sites in the precursor for the sea
A:Reference number: A44308; MUID:93054550; PMID:1429603
A:Accession: A44308
A:Status: preliminary
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-435 <SCH>
A:Cross-references: UNIPROT:P10419; GB:M98269; NID:G155702; PIDN:AA27738.1; PID:G155703
A:Note: sequence extracted from NCBI backbone (NCBIN:117102, NCBI:117104)
C:Keywords: neuropeptide

Query Match 52.6%; Score 60; DB 2; Length 435;
Best Local Similarity 35.7%; Pred. No. 20;
Matches 10; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOEKXXXXXXXXXXWSHPQEK 28
DB 66 FSDPQFWKGRFSDPQFWKGRFSDPQFWK 93

RESULT 2
T25450
hypothetical protein B0412.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25450
R:Bentley, D.
Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid B0412.
A:Reference number: Z20037
A:Accession: T25450
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1107 <BEN>
A:Cross-references: UNIPROT:P90984; EMBL:U80953; PIDN:AA852556.1; GSPDB:GN00021; CESP:B0
A:Experimental source: strain Bristol N2; clone B0412
C:Genetics:
A:Gene: CESP:B0412.3
A:Map position: 3
A:Introns: 39/2, 70/1, 133/2, 288/3, 386/1, 470/3, 568/3, 631/2, 733/2, 819/3, 931/1, 96
C:Superfamily: Caenorhabditis elegans hypothetical protein B0412.3

Query Match 52.6%; Score 60; DB 2; Length 1107;
Best Local Similarity 21.4%; Pred. No. 60;
Matches 6; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQEK 28
 DB 76 WFOKLDKVPALIVFIDLEWDPMSWP 103

RESULT 3

probable transcription factor - fission yeast (schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 R/Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 A/Submitted to the EMBL Data Library, September 1998
 A/Reference number: Z17313
 A/Accession: T11677
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-601 <SEE>
 A/Cross-references: UNIPROT:O74345; EMBL:AL031536; NID:el319499; PID:el319504
 C/Genetics:
 A/Map position: IIR
 A/Intons: 39/1; 74/3; 106/2
 A/Note: SPBC21D10.05c

Query Match 52.2%; Score 59.5; DB 2; Length 601;
 Best Local Similarity 31.0%; Pred. No. 35;
 Matches 9; Conservative 13; Mismatches 2; Indels 5; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXWSHP 24
 DB 68 WSNQLEKMKMGWINANRYMNPPLSP 96

RESULT 4

conserved hypothetical protein SPCC1281.07c - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T40926
 R/Volckaert, G.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 A/Submitted to the EMBL Data Library, January 1999
 A/Reference number: Z21957
 A/Accession: T40926
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-313 <VOL>
 A/Cross-references: UNIPROT:O94524; EMBL:AL035218; PIDN:CAA22828.1; GSPDB:GN00068; SPDB:
 A/Experimental source: strain 972h-; cosmid c1281
 C/Genetics:
 A/Gene: SPDB:SPCC1281.07c
 A/Map position: 3
 C/Superfamily: glutathione S-transferase

Query Match 50.9%; Score 58; DB 2; Length 313;
 Best Local Similarity 28.0%; Pred. No. 26;
 Matches 7; Conservative 13; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQ 25
 DB 271 WGRPAFHETDPMKIKCHYTQSHQ 295

RESULT 5

hypothetical protein ZK1248.15 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
 C/Accession: T34497
 R/Latelle, P.
 A/Submitted to the EMBL Data Library, June 1995
 A/Description: The sequence of C. elegans cosmid ZK1248.

A/Reference number: Z21534
 A/Accession: T34497
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-652 <LAT>
 A/Cross-references: UNIPROT:Q23425; EMBL:U29244; PIDN:AACT1096.1; GSPDB:GN00020; CESP:ZT
 A/Experimental source: strain Bristol N2; clone ZK1248
 C/Genetics:
 A/Gene: CESP:ZK1248.15
 A/Map position: 2
 A/Intons: 29/2; 107/3; 172/2; 264/3; 310/3; 360/1; 443/1; 575/3
 C/Superfamily: Caenorhabditis elegans hypothetical protein ZK1248.15; WW repeat homology
 F:198-236/Domain: WW repeat homology <WWR>

Query Match 50.9%; Score 58; DB 2; Length 652;
 Best Local Similarity 21.4%; Pred. No. 61;
 Matches 6; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQEK 28
 DB 207 WWHTEKKKKFYNDKXESLMDHNTTRK 234

RESULT 6

hypothetical protein At2g41600 [imported] - Arabidopsis thaliana
 N/Alternate names: hypothetical protein T3266.12
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C/Accession: T00816; G84843
 R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
 A/Submitted to the EMBL Data Library, November 1997
 A/Description: Arabidopsis thaliana chromosome II BAC T3266 genomic sequence.
 A/Reference number: Z14163
 A/Accession: T00816
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-151 <ROU>

A/Cross-references: UNIPROT:O22222; EMBL:AC002510; NID:G2618683; PID:G2618695
 A/Experimental source: cultivar Columbia
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Taiton, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: A84420; WUID:20083487; PMID:10617197
 A/Accession: G84843
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-151 <STO>
 A/Cross-references: GB:AE002093; NID:G2618695; PIDN:AAB84342.1; GSPDB:GN00139
 C/Genetics:
 A/Gene: T3266.12; At2g41600
 A/Map position: 2
 A/Intons: 38/3
 C/Superfamily: Arabidopsis thaliana hypothetical protein At2g41600

Query Match 50.0%; Score 57; DB 2; Length 151;
 Best Local Similarity 25.0%; Pred. No. 15;
 Matches 6; Conservative 15; Mismatches 3; Indels 0; Gaps 0;

QY 2 SHPOFEKXXXXXXXXXXWSHPQ 25
 DB 33 SHRFQGVETGSLGPKLMDSPF 56

RESULT 7

hypothetical protein - Synechocystis sp. (strain PCC 6803)
 C/Species: Synechocystis sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C/Accession: S76147

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimizu, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-135, 1996
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
B.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76147
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <KAN>
A:Cross-references: UNIPROT:E74312; EMBL:D90914; GB:AB001339; NID:g1653477; PIDN:BA01840
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match
Best Local Similarity 50.0%; Score 57; DB 2; Length 206;
Matches 6; Conservative 16; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SHPOEKXXXXXXXXXXSHPO 25
|||:|||||:|||||:|||||:
Db 58 SHPKVQELLELSQWSEGVWSSPE 81

RESULT 8
A:Accession: A45462
C:4.2.3.9 aristochochene synthase () - Penicillium roqueforti
C:Species: Penicillium roqueforti
C>Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: A45462
R:Proctor, R.H.; Hohn, T.M.
J. Biol. Chem. 268, 4543-4548, 1993
A>Title: Aristochochene synthase, isolation, characterization, and bacterial expression
A:Reference number: A45462; MUID:93179472; PMID:8440737
A:Accession: A45462
A:Molecule type: DNA
A:Residues: 1-342 <PRO>
A:Cross-references: UNIPROT:O03471; GB:L05193; NID:g169177; PIDN:AA33694.1; PID:g169178
A>Note: sequence extracted from NCBI backbone (NCBI:125996, NCBI:125997)
C:Comment: This enzyme is a sesquiterpene cyclase involved in synthesis of the sesquiterpene
C:Genetics:
A:Gene: At11
C:Superfamily: aristochochene synthase
C:Keywords: transferase

Query Match
Best Local Similarity 49.6%; Score 56.5; DB 2; Length 342;
Matches 7; Conservative 16; Mismatches 4; Indels 3; Gaps 1;

Qy 1 WS---HPOEKXXXXXXXXXXSHPO 27
|||:|||||:|||||:|||||:
Db 46 WSYLCHPRVKVQDEVGDFLENMKRPFK 75

RESULT 9
E90973
hypothetical protein EC62757 [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90973
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
A:Reference number: A9629; MUID:21156231; PMID:11258796
A:Accession: E90973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAV>
A:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BA036180.1; PID:g13362225; GSPDB:C
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC62757

Query Match
49.1%; Score 56; DB 2; Length 103;
Matches 5; Conservative 17; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 22.7%; Pred. No. 14;
Matches 5; Conservative 17; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSHPOEKXXXXXXXXXXMS 22
|||:|||||:|||||:|||||:
Db 17 WTHPEYKFCGDEYISTEEFN 38

RESULT 10
A:Accession: A85821
unknown protein encoded within prophage CP-9330 [imported] - *Escherichia coli* (strain O1
C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85821
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Linn, A.; Diallanata, E.; Potamoultis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85821
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:Q8X4V0; GB:AE005174; NID:g12516136; PIDN:AA057029.1; GSPDB:C
A:Experimental source: strain O157:H7, substrain ED933
C:Genetics:
A:Gene: Z3120

Query Match
Best Local Similarity 49.1%; Score 56; DB 2; Length 103;
Matches 5; Conservative 17; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSHPOEKXXXXXXXXXXMS 22
|||:|||||:|||||:|||||:
Db 17 WTHPEYKFCGDEYISTEEFN 38

RESULT 11
C85713
unknown protein encoded within prophage CP-9330 [imported] - *Escherichia coli* (strain O1
C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: C85713
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Linn, A.; Diallanata, E.; Potamoultis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85713
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:g12515046; PIDN:AA056167.1; GSPDB:C
A:Experimental source: strain O157:H7, substrain ED933
C:Genetics:
A:Gene: Z2097

Query Match
Best Local Similarity 49.1%; Score 56; DB 2; Length 103;
Matches 6; Conservative 15; Mismatches 1; Indels 0; Gaps 0;

Qy 7 ECKXXXXXXXXXXSHPOEK 28
|||:|||||:|||||:|||||:
Db 3 EIKEMPVARDGYWTHPEYK 24

RESULT 12
A90904
hypothetical protein EC62201 [imported] - *Escherichia coli* (strain O157:H7, substrain RI
C:Species: *Escherichia coli*
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: A90904
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gend
 A:Reference number: A59629; MUID:21156231; PMID:11258796
 A:Accession: A5904
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-103 <HAY>
 A:Cross-references: UNIPROT:O8X549; GB:BA000007; PIDN:BA35624.1; PID:G13361667; GSPDB:G
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: EC62201

Query Match 49.1%; Score 56; DB 2; Length 103;
 Best Local Similarity 27.3%; Pred. No. 14;
 Matches 6; Conservative 15; Mismatches 1; Indels 0; Gaps 0;
 Oy 7 EKXXXXXXXXXXXXXSHPOFEK 28
 Db 3 EIKEMPVVRDGYGWTPEYER 24

RESULT 13
 A39172
 Antho-Ramide neuropeptide 19 repeat precursor - sea anemone (*Calliactis parasitica*)
 C:Species: *Calliactis parasitica*
 C>Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 09-Jul-2004
 C:Accession: A39172
 R:Darmer, D.; Schmutzler, C.; Diekhoff, D.; Grimmelikhuijzen, C.J.P.
 Proc. Natl. Acad. Sci. U.S.A. 88, 2555-2559, 1991
 A>Title: Primary structure of the precursor for the sea anemone neuropeptide Antho-RFamide
 A:Reference number: A39172; MUID:91172845; PMID:1706527
 A:Accession: A39172
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-334 <DAR>
 A:Cross-references: UNIPROT:Q01133; GB:M59166; NID:G156133; PIDN:AAA27878.1; PID:G156134
 C:Keywords: neuropeptide

Query Match 49.1%; Score 56; DB 2; Length 334;
 Best Local Similarity 36.0%; Pred. No. 53;
 Matches 9; Conservative 13; Mismatches 3; Indels 0; Gaps 0;
 Oy 4 POPEXXXXXXXXXXXXXSHPOFEK 28
 Db 56 PQFWKGRFSDPQFWKGRFSDPQFWK 80

RESULT 14
 AD3492
 acetate-CoA ligase (EC 6.2.1.16) [imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
 C:Accession: AD3492
 R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AD3492
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-335 <KUR>
 A:Cross-references: UNIPROT:O8YEF6; GB:AB009917; PIDN:AAL53103.1; PID:G17983968; GSPDB:G
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BME11922
 A:Map position: I
 C:Keywords: acid-thiol ligase; coenzyme A

Query Match 49.1%; Score 56; DB 2; Length 335;
 Best Local Similarity 26.9%; Pred. No. 53;
 Matches 7; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

Oy 1 WSHPOPEXXXXXXXXXXXXXSHPOF 26
 Db 164 WNDAGQEKYQAYFPERFDNWHGDF 189

RESULT 15
 S30014
 hypothetical protein YKL012w - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein YKL165
 C:Species: *Saccharomyces cerevisiae*
 C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 09-Jul-2004
 C:Accession: S30014; S37823; S37825
 R:Pascolo, S.; Ghazvini, M.; Boyer, J.; Collea, L.; Thierry, A.; Dujon, B.
 Yeast 8, 987-995, 1992
 A>Title: The sequence of a 9.3 kb segment located on the left arm of the yeast chromosome
 A:Reference number: S30013; MUID:93127732; PMID:1481574
 A:Accession: S30014
 A:Molecule type: DNA
 A:Residues: 1-583 <PAS>
 A:Cross-references: UNIPROT:P33203; GB:S53418; NID:G263497; PIDN:AMB24902.1; PID:G263498
 A:Experimental source: strain S288C
 R:Boyer, J.; Pascolo, S.; Richard, G.F.; Ghazvini, M.; Collea, L.; Thierry, A.; Monne
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S37813
 A:Accession: S37823
 A:Molecule type: DNA
 A:Residues: 1-583 <BOY>
 A:Cross-references: EMBL:Z28012; NID:G485996; PIDN:CAA81847.1; PID:G485997; MIPS:YKL012w
 A:Experimental source: strain S288C
 R:Wiemann, S.; Voss, H.; Schwaiger, C.; Rupp, T.; Grothues, D.; Sengen, C.; Stegemann, J.
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S37825
 A:Accession: S37825
 A:Molecule type: DNA
 A:Residues: 1-583 <WIE>
 A:Cross-references: EMBL:Z28012; NID:G485996; PIDN:CAA81847.1; PID:G485997; MIPS:YKL012w
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:PPP40
 A:Cross-references: SGD:S0001495; MIPS:YKL012w
 A:Map position: 11L
 F:1-35/Domain: WW repeat homology #status atypical <WW1>
 F:39-76/Domain: WW repeat homology <WW2>

Query Match 49.1%; Score 56; DB 2; Length 583;
 Best Local Similarity 32.0%; Pred. No. 16+02;
 Matches 8; Conservative 13; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 WSHPOPEXXXXXXXXXXXXXSHPOF 25
 Db 67 WTPAEKKEVPEIAEQKDTVSHAQ 91

Search completed: March 2, 2005, 12:29:02
 Job time : 22.3984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 104.033 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ16
Perfect score: 114
Sequence: 1 wshpgefekxxxxxxxxxxwshpgefek 28

Scoring table: BLOSUM62DX
gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	67	58.8	1749	2	Q8T0W6	Q8T0W6 echinococcus
2	64	56.1	704	2	Q8CJD2	Q8CJD2 rattus norvegicus
3	63	55.3	743	2	Q8CH90	Q8CH90 rattus norvegicus
4	61	53.5	304	2	Q74JBE	Q74JBE lactobacillus
5	61	53.5	869	2	Q62MT8	Q62MT8 burkholderia
6	61	53.5	869	2	Q63Y50	Q63Y50 burkholderia
7	60.5	53.1	107	2	Q6Z315	Q6Z315 oryza sativa
8	60.5	53.1	719	2	Q87NM4	Q87NM4 vibrio parahaemolyticus
9	60	52.6	429	1	FMK2 ANTEL	P10494 anthopleura
10	60	52.6	435	1	FMK1 ANTEL	P10419 anthopleura
11	60	52.6	506	2	Q73TF5	Q73TF5 mycobacterium
12	60	52.6	508	1	MATK MARSC	Q8WK9 maritimum s
13	60	52.6	508	2	Q9B8G0	Q9B8G0 vanovaynell
14	60	52.6	508	2	Q9B8G3	Q9B8G3 oserya coulteri
15	60	52.6	650	2	Q83DG9	Q83DG9 coxiella burnetii
16	60	52.6	671	2	Q6YQD6	Q6YQD6 onion yellows
17	60	52.6	1015	2	Q6C664	Q6C664 yarrowia lipolytica
18	60	52.6	1107	2	P90984	P90984 caenorhabditis
19	59.5	52.2	601	2	Q74345	Q74345 schizosaccharomyces
20	59	51.8	759	2	Q8RYC2	Q8RYC2 oryza sativa
21	59	51.8	2333	2	Q8RSP8	Q8RSP8 candida glabrata
22	58	50.9	166	2	Q8NYE4	Q8NYE4 staphylococcus
23	58	50.9	166	2	Q6GCH6	Q6GCH6 staphylococcus
24	58	50.9	297	2	Q891J0	Q891J0 bradyrhizobium
25	58	50.9	313	2	Q94524	Q94524 schizosaccharomyces
26	58	50.9	337	2	Q6VY41	Q6VY41 vibrio parahaemolyticus
27	58	50.9	364	2	Q6FTJ1	Q6FTJ1 candida glabrata
28	58	50.9	431	2	Q7N8A2	Q7N8A2 photorhabdus
29	58	50.9	471	2	Q23425	Q23425 caenorhabditis
30	57	50.0	59	2	Q756K2	Q756K2 neurospora
31	57	50.0	151	2	Q22222	Q22222 arabidopsis

32	57	50.0	192	2	Q6UNF7 oryza sativ
33	57	50.0	206	2	P74312 synechocyst
34	57	50.0	247	2	Q82S44 nitrosomona
35	57	50.0	305	2	Q8W403 oryza sativ
36	57	50.0	305	2	Q6Z0Y9 oryza sativ
37	57	50.0	355	2	Q9VX10 dirosophila
38	57	50.0	541	2	Q987P4 rhizobium l
39	57	50.0	590	2	Q6CVV1 kluyveromyc
40	57	50.0	652	2	Q988Y6 rhizobium l
41	57	50.0	719	1	GLB2_XANAC
42	57	50.0	1698	1	CUL7_HUMAN
43	56.5	49.6	144	2	Q8ZG44 streptomyce
44	56.5	49.6	224	2	Q8RYV8 coffea arab
45	56.5	49.6	322	2	Q6CSW9 kluyveromyc

ALIGNMENTS

RESULT 1	ID	Q8T0W6	PRELIMINARY;	PRT; 1749 AA.
AC	Q8T0W6			
DT	01-JUN-2002 (TRMBLrel. 21, Created)			
DT	01-JUN-2002 (TRMBLrel. 21, Last sequence update)			
DT	01-MAR-2004 (TRMBLrel. 26, Last annotation update)			
DE	Putative insulin receptor precursor.			
GN	Name=Irh;			
OS	Echinococcus multilocularis.			
OC	Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;			
OC	Cyclophyllidae; Taeniidae; Echinococcus.			
OX	NCBI_Taxid=6211;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22557731; PubMed=12670515; DOI=10.1016/S0020-7519(02)00265-5;			
RA	Konrad C., Kroner A., Spiliotis M., Zavala-Gongora R., Brehm K.;			
RT	"Identification and molecular characterisation of a gene encoding a			
RT	member of the insulin receptor family in Echinococcus			
RT	multilocularis."			
RL	Int. J. Parasitol. 33:301-312(2003).			
DR	EMBL; AJ458426; CAD30260.1; -			
DR	HSSP; P08069; 1JOH.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0005006; F:epidermal growth factor receptor activity; IEA.			
DR	GO; GO:0004872; F:receptor activity; IEA.			
DR	GO; GO:0016740; F:transferase activity; IEA.			
DR	GO; GO:0006468; P:protein amino acid phosphorylation; IEA.			
DR	GO; GO:0007159; P:transmembrane receptor protein tyrosine kin. . .; IEA.			
DR	InterPro; IPR00494; EGFR_L.			
DR	InterPro; IPR003961; FN_III.			
DR	InterPro; IPR008957; FN_III-like.			
DR	InterPro; IPR006211; Furin-like.			
DR	InterPro; IPR006212; Furin repeat.			
DR	InterPro; IPR009030; Grow_Fac_recept.			
DR	InterPro; IPR000585; Hemopexin.			
DR	InterPro; IPR01009; Kinase-like.			
DR	InterPro; IPR000719; Prot_Kinase.			
DR	InterPro; IPR001245; Tyr_kinase.			
DR	InterPro; IPR008266; Tyr_kinase_AS.			
DR	Pfam; PF00757; Furin-like; 1.			
DR	Pfam; PF01030; Recep_L_domain; 1.			
DR	PRINTS; PR00109; TYRKINASE.			
DR	SMART; SMO0060; FN3; 3.			
DR	SMART; SMO0261; FU; 1.			
DR	SMART; SMO0219; TYRK; 1.			
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.			
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
KW	Kinase; Receptor; Signal; Transferase; Tyrosine-protein kinase.			
FT	SIGNAL 1 32 Potential.			
FT	CHAIN 33 1749 Potential.			
SQ	SEQUENCE 1749 AA; 192538 MW; 9A8308BFE9597DA CRC64;			

Query Match 58.8%; Score 67; DB 2; Length 1749;
 Best Local Similarity 32.0%; Pred. No. 53;
 Matches 8; Conservative 15; Mismatches 2; Indels 0; Gaps 0;

RESULT 2

Q8CUD2 PRELIMINARY; PRT; 704 AA.

AC Q8CUD2; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 OS Guanylyl cyclase alpha 1 subunit.
 OC Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura I., Suzuki N.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB096020; BAC24016.1; -
 DR HSSP; P30803; IAZS.
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR01054; G_cyclase.
 DR Pfam; PF00211; Guanylate_cyc; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS0125; GUANYLATE_CYCLASES_2; 1.
 KW Lyase.
 SQ SEQUENCE 704 AA; 79188 MW; F73FC97B685456A CRC64;

RESULT 3

Q8CH90 PRELIMINARY; PRT; 743 AA.

AC Q8CH90; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 OS Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
 OC Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura I., Yao Y., Suzuki N.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB097860; BAC44887.1; -
 DR HSSP; P30803; IAZS.
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR InterPro; IPR01054; G_cyclase.
 DR InterPro; IPR009080; tRNA_syn_1a_bind.
 DR Pfam; PF00211; Guanylate_cyc; 1.

DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS0125; GUANYLATE_CYCLASES_2; 1.
 KW Lyase.
 SQ SEQUENCE 743 AA; 83251 MW; ACP5CS3E0982813A CRC64;

Query Match 55.3%; Score 63; DB 2; Length 743;
 Best Local Similarity 40.0%; Pred. No. 70;
 Matches 8; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

Q74JB6 PRELIMINARY; PRT; 304 AA.

AC Q74JB6; 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 OS Hypothetical protein.
 GN Ordered locus names=L01161;
 OC Lactobacillus johnsonii.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 CX NCBI_TaxId=33959;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=NC 533;
 RC PubMed14966310;
 RA Fridmore R.D., Berger B., Desiere F., Vilanova D., Barretto C.,
 RA Pitter A.-C., Zwaalen M.-C., Rouvet M., Altemann E., Barrangu R.,
 RA Mollet B., Mercier A., Kleenhammer T., Arigoni F., Schell W.A.;
 RT "The genome sequence of the probiotic intestinal bacterium
 RT Lactobacillus johnsonii NC 533".
 RL Proc. Natl. Acad. Sci. U.S.A. 101:2512-2517(2004).
 DR EMBL; AB017203; AAS08983.1; -
 DR GO; GO:0001760; F:aminocarboxymuconate-semialdehyde decarboxy. . .; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR02160; Prot inh kunz-19.
 DR PROSITE; PS00283; SOYBEAN_KUNITZ; UNKNOWN_1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 304 AA; 34432 MW; A81B9DC23410B5A4 CRC64;

Query Match 53.5%; Score 61; DB 2; Length 304;
 Best Local Similarity 29.2%; Pred. No. 46;
 Matches 7; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Q74JB6 PRELIMINARY; PRT; 869 AA.

AC Q62MT8; 25-OCT-2004 (TREMBlrel. 28, Created)
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
 DE 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
 OS DNA topoisomerase IIL (EC 5.99.1.2).
 GN Name=topB; ORFNames=BMA0139.
 OS Burkholderia mallei ATCC 23344.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia.
 CX NCBI_TaxId=243160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=ATCC 23344;
 RA Nierman W.C., Deshazer D., Kim H.S., Tetteijn H., Nelson K.E.,
 RA Feldblyum T., Ulrich R.L., Ronning C.M., Brinkac L.M., Daugherty S.C.,

```
RA Davidson T.D., Deboy R.T., Dimitrov G., Dodson R.J., Durkin A.S.,
RA Gwinn M.L., Haft D.H., Khouri H., Kolonay J.F., Madupu R.,
RA Mohammad Y., Nelson W.C., Radune D., Romero C.M., Sarría S.,
RA Selengut U., Shandlin C., Sullivan S.A., White O., Yu Y., Zatar N.,
RA Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251 (2004).
DR EMBL; CP000010; AAU48686.1; -.
KW Isoenzyme.
SQ SEQUENCE 869 AA; 96949 MW; 32EB90771AA5E556 CRC64;

Query Match 53.5%; Score 61; DB 2; Length 869;
Best Local Similarity 29.6%; Pred. No. 1.6e+02;
Matches 8; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXXWSHPQFE 27
Db 237 WDPKFKKDEFPKRDRLMLPAAE 263

RESULT 6
Q63YS0 PRELIMINARY; PRT; 869 AA.
AC Q63YS0;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DE Putative DNA topoisomerase III (EC 5.99.1.2).
GN ORFNames=BP5L0118;
OS Burkholderia pseudomallei K96243.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxId=272560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K96243;
RX PubMed=15377794;
RA Holden M.T.G., Tilball R.W., Peacock S.J., Cerdano-Tarraga A.M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.,
RA Bentley S.D., Sebahia M., Thomson N.R., Baeson N., Beacham I.R.,
RA Brockle K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosser B., Davis P., Deshazer D.,
RA Felwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagsels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinovitch E., Rutherford K., Sanders M., Simmonds M.,
RA Songeivylá S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrett B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245 (2004).
DR EMBL; BX571965; CAH34104.1; -.
KW Isoenzyme.
SQ SEQUENCE 869 AA; 96945 MW; 32ED56B71AA5E152 CRC64;

Query Match 53.5%; Score 61; DB 2; Length 869;
Best Local Similarity 29.6%; Pred. No. 1.6e+02;
Matches 8; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXXWSHPQFE 27
Db 237 WDPKFKKDEFPKRDRLMLPAAE 263

RESULT 7
Q62315 PRELIMINARY; PRT; 107 AA.
AC Q62315;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE DCL protein-like.
GN Name=OJ1004 AII.16-2; Synonyms=P0539D10.35-2;
OS Oryza sativa (japonica cultivar-group).
```

```
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhacroidae; Oryzaceae; Oryza.
OX NCBI_TaxId=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005287; BAD17315.1; -.
DR EMBL; AP004817; BAD17127.1; -.
SQ SEQUENCE 107 AA; 12839 MW; 00F06AF3A12B0EB1 CRC64;

Query Match 53.1%; Score 60.5; DB 2; Length 107;
Best Local Similarity 36.0%; Pred. No. 16;
Matches 9; Conservative 14; Mismatches 1; Indels 1; Gaps 1;

Qy 3 HPQFEKXXXXXXXXXXWSHPQFE 27
Db 30 HPQFEKXIGCGIDYLVG-LHPEFE 53

RESULT 8
Q87NW4 PRELIMINARY; PRT; 719 AA.
AC Q87NW4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein VP1754.
GN OrderedLocNames=VP1754;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxId=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749 (2003).
DR EMBL; AP005079; BAC60017.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0007165; F:signal transduction; IEA.
DR InterPro; IPR001633; EAL.
DR InterPro; IPR001630; GGDEF.
DR InterPro; IPR003660; His_kin_HAMP.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GGDEF; 1.
DR Pfam; PF0672; HAMP; 1.
DR SMART; SM00267; DUF1.1.
DR SMART; SM00052; DUF2.1.
DR SMART; SM00304; HAMP; 1.
DR TIGRfams; TIGR00254; GGDEF; 1.
DR PROSITE; PS50883; EAL; 1.
DR PROSITE; PS50887; GGDEF; 1.
DR PROSITE; PS50885; HAMP; 1.
KW Complete proteome.
SQ SEQUENCE 719 AA; 81471 MW; CAB9568F0C0C6F1 CRC64;

Query Match 53.1%; Score 60.5; DB 2; Length 719;
Best Local Similarity 26.9%; Pred. No. 1.5e+02;
Matches 7; Conservative 16; Mismatches 2; Indels 1; Gaps 1;

Qy 1 WSHPOFE-KXXXXXXXXXXWSHPQ 25
```

Db 473 WYQPKEDLKTQYGVGEALVRNHP 498

RESULT 9

ID	FMRL ANTEL	STANDARD;	PRT;	429 AA.
AC	016994;			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUN-2004 (Rel. 44, Last annotation update)			
DE	Antho-Rfamide, neuropeptides type 2 precursor.			
OS	Anthopleura elegantissima (Sea anemone).			
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;			
OC	Nyctanthae; Actiniidae; Anthopleura.			
OX	NCBI_TaxID=6110;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93054550; PubMed=1429603;			
RA	Schmutzler C., Darmer D., Diekhoff D., Grimmelikhuijzen C.J.P.;			
RT	"Identification of a novel type of processing sites in the precursor for the sea anemone neuropeptide Antho-Rfamide (<Glu-Gly-Arg-Phe-NH2) from Anthopleura elegantissima."			
RL	J. Biol. Chem. 267:22534-22541 (1992).			
CC	[2]			
CC	PARTIAL SEQUENCE (ANTHO-RFAMIDE).			
CC	MEDLINE=87092339; PubMed=2879288;			
CC	Grimmelikhuijzen C.J.P., Graff D.;			
CC	"Isolation of pyroglu-Gly-Arg-Phe-NH2 (Antho-Rfamide), a neuropeptide from sea anemones."			
CC	Proc. Natl. Acad. Sci. U.S.A. 83:9817-9821 (1986).			
CC	-1- FUNCTION: Not known but it could act as a transmitter at neuromuscular synapses.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide) family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M99170; AAA27738.1; -.			
DR	InterPro; IPR005544; FARP.			
DR	Pfam; PF01581; FARP; 14.			
KW	Annotation: Direct protein sequencing; Neuropeptide; Repeat; Signal.			
FT	SIGNAL	1	22	Potential.
FT	PEPTIDE	234	237	Antho-Rfamide.
FT	PEPTIDE	242	245	Antho-Rfamide.
FT	PEPTIDE	250	253	Antho-Rfamide.
FT	PEPTIDE	258	261	Antho-Rfamide.
FT	PEPTIDE	266	269	Antho-Rfamide.
FT	PEPTIDE	274	277	Antho-Rfamide.
FT	PEPTIDE	290	293	Antho-Rfamide.
FT	PEPTIDE	298	301	Antho-Rfamide.
FT	PEPTIDE	306	309	Antho-Rfamide.
FT	PEPTIDE	322	325	Antho-Rfamide.
FT	PEPTIDE	330	333	Antho-Rfamide.
FT	PEPTIDE	343	346	Antho-Rfamide.
FT	PEPTIDE	356	359	Antho-Rfamide.
FT	PEPTIDE	369	372	Antho-Rfamide.
FT	MOD_RES	237	237	Phenylalanine amide (G-238 provides amide group).
FT	MOD_RES	245	245	Phenylalanine amide (G-246 provides amide group).
FT	MOD_RES	253	253	Phenylalanine amide (G-254 provides amide group).
FT	MOD_RES	261	261	Phenylalanine amide (G-262 provides amide group).
FT	MOD_RES	269	269	Phenylalanine amide (G-270 provides amide group).

FT MOD_RES 277 277 (group).

FT MOD_RES 293 293 Phenylalanine amide (G-294 provides amide group).

FT MOD_RES 301 301 Phenylalanine amide (G-302 provides amide group).

FT MOD_RES 309 309 Phenylalanine amide (G-310 provides amide group).

FT MOD_RES 325 325 Phenylalanine amide (G-326 provides amide group).

FT MOD_RES 333 333 Phenylalanine amide (G-334 provides amide group).

FT MOD_RES 346 346 Phenylalanine amide (G-347 provides amide group).

FT MOD_RES 359 359 Phenylalanine amide (G-360 provides amide group).

FT MOD_RES 372 372 Phenylalanine amide (G-373 provides amide group).

SO SEQUENCE 429 AA; 50564 MW; 7C54F5C606D537F4 CRC64;

Query Match

Best Local Similarity 52.6%; Score 60; DB 1; Length 429;

Matches 10; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQFK 28

Db 66 FSDPQFMKGRFSDPQFMKGRFSDPQFK 93

RESULT 10

ID	FMRL ANTEL	STANDARD;	PRT;	435 AA.
AC	P10419;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	05-JUN-2004 (Rel. 44, Last annotation update)			
DE	Antho-Rfamide, neuropeptides type 1 precursor.			
OS	Anthopleura elegantissima (Sea anemone).			
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;			
OC	Nyctanthae; Actiniidae; Anthopleura.			
OX	NCBI_TaxID=6110;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93054550; PubMed=1429603;			
RA	Schmutzler C., Darmer D., Diekhoff D., Grimmelikhuijzen C.J.P.;			
RT	"Identification of a novel type of processing sites in the precursor for the sea anemone neuropeptide Antho-Rfamide (<Glu-Gly-Arg-Phe-NH2) from Anthopleura elegantissima."			
RL	J. Biol. Chem. 267:22534-22541 (1992).			
CC	[2]			
CC	PARTIAL SEQUENCE (ANTHO-RFAMIDE).			
CC	MEDLINE=87092339; PubMed=2879288;			
CC	Grimmelikhuijzen C.J.P., Graff D.;			
CC	"Isolation of pyroglu-Gly-Arg-Phe-NH2 (Antho-Rfamide), a neuropeptide from sea anemones."			
CC	Proc. Natl. Acad. Sci. U.S.A. 83:9817-9821 (1986).			
CC	-1- FUNCTION: Not known but it could act as a transmitter at neuromuscular synapses.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide) family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M98269; AAA27738.1; -.			
DR	PIR; A26666; ECXAA.			

DR PIR; A44308; A44308.
 DR InterPro; IPR002544; PARP.
 DR Pfam; PF01581; PARP; 13.
 KW Amidation; Direct protein sequencing; Neuropeptide;
 KW Pyroglutamate; Pyroglutamate; Pyroglutamate; Pyroglutamate;
 FT SIGNAL 1 22
 FT PEPTIDE 194 197 Antho-Rfam.
 FT PEPTIDE 202 205 Antho-Rfam.
 FT PEPTIDE 210 213 Antho-Rfam.
 FT PEPTIDE 218 221 Antho-Rfam.
 FT PEPTIDE 226 229 Antho-Rfam.
 FT PEPTIDE 234 237 Antho-Rfam.
 FT PEPTIDE 242 245 Antho-Rfam.
 FT PEPTIDE 250 253 Antho-Rfam.
 FT PEPTIDE 258 261 Antho-Rfam.
 FT PEPTIDE 266 269 Antho-Rfam.
 FT PEPTIDE 274 277 Antho-Rfam.
 FT PEPTIDE 282 285 Antho-Rfam.
 FT PEPTIDE 290 293 Antho-Rfam.
 FT PEPTIDE 298 301 Antho-Rfam.
 FT PEPTIDE 306 309 Antho-Rfam.
 FT PEPTIDE 314 317 Antho-Rfam.
 FT PEPTIDE 322 325 Antho-Rfam.
 FT PEPTIDE 330 333 Antho-Rfam.
 FT PEPTIDE 336 339 Antho-Rfam.
 FT PEPTIDE 343 346 Antho-Rfam.
 FT PEPTIDE 356 359 Antho-Rfam.
 FT PEPTIDE 369 372 Antho-Rfam.
 FT DOMIN 376 386 Antho-Rfam.
 FT MOD_RES 194 197 Antho-Rfam.
 FT MOD_RES 197 199 Antho-Rfam.
 FT MOD_RES 205 205 Antho-Rfam.
 FT MOD_RES 213 213 Antho-Rfam.
 FT MOD_RES 221 221 Antho-Rfam.
 FT MOD_RES 229 229 Antho-Rfam.
 FT MOD_RES 237 237 Antho-Rfam.
 FT MOD_RES 245 245 Antho-Rfam.
 FT MOD_RES 253 253 Antho-Rfam.
 FT MOD_RES 261 261 Antho-Rfam.
 FT MOD_RES 269 269 Antho-Rfam.
 FT MOD_RES 277 277 Antho-Rfam.
 FT MOD_RES 285 285 Antho-Rfam.
 FT MOD_RES 293 293 Antho-Rfam.
 FT MOD_RES 301 301 Antho-Rfam.
 FT MOD_RES 309 309 Antho-Rfam.
 FT MOD_RES 317 317 Antho-Rfam.
 FT MOD_RES 325 325 Antho-Rfam.
 FT MOD_RES 333 333 Antho-Rfam.
 FT MOD_RES 346 346 Antho-Rfam.
 FT MOD_RES 359 359 Antho-Rfam.
 FT MOD_RES 372 372 Antho-Rfam.
 FT SEQUENCE 435 AA; 50940 MW; B0C44020CD58D61 CRC64;

Query Match 52.6%; Score 60; DB 1; Length 435;
 Best Local Similarity 35.7%; Pred. No. 97;
 Matches 10; Conservative 14; Mismatches 4; Indels 0; Gaps 0;
 QY 1 WSHPEKXXXXXXXXXXXXXWSHPQFEK 28
 DB 66 FSDPQFMKGRFSDPQFMKGRFSDPQFMK 93
 RESULT 11
 ID Q73TF5 PRELIMINARY; PRT; 506 AA.
 AC Q73TF5;
 DT 05-JUL-2004 (TRENBLREL. 27, Created)
 DT 05-JUL-2004 (TRENBLREL. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLREL. 27, Last annotation update)
 DE Pap3.2.
 GN Name=pap3; OrderedLocustNames=MAP3763C;
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=k10;
 RA Li L., Bannantine J., Zhang Q., Ammons A., Alt D., Kapur V.;
 RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB017240; AAS06313.1; .
 KW Complete proteome.
 SQ SEQUENCE 506 AA; 56048 MW; 9E736B347E19153D CRC64;
 Query Match 52.6%; Score 60; DB 2; Length 506;
 Best Local Similarity 31.8%; Pred. No. 12e+02;
 Matches 7; Conservative 12; Mismatches 3; Indels 0; Gaps 0;
 QY 3 HPQFEKXXXXXXXXXXXXXWSHP 24
 DB 476 HPQGSREARREARREARREARREARRE 497
 RESULT 12
 MATR MARSC STANDARD; PRT; 508 AA.
 ID Q8WK9;
 AC Q8WK9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Maturase K (Intron maturase).
 GN Name=matk;
 OS Marathrum schiedeanum.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Malpighiales; Podostemaceae; Marathrum.
 OX NCBI_TaxID=116737;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kita Y., Kato M.;
 RT Phylogenetic relationships of the aquatic angiosperm family
 RT Podostemaceae inferred from matk sequence data.
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Probably assists in splicing chloroplast group II
 CC introns (by similarity).
 CC -1- SIMILARITY: Belongs to the intron maturase family 2. Matk
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: AB038195; BAB3156.1; -
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
SQ SEQUENCE 508 AA; 61342 MW; ECCFSB416B0AB5C3 CRC64;

Query Match
Best Local Similarity 52.6%; Score 60; DB 1; Length 508;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPOFEK 28
Db 168 WLKDPASLHFLRLEFHEYMNWSPSEK 195

RESULT 13
O9BBG0 PRELIMINARY; PRT; 508 AA.
AC O9BBG3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Maturase K.
DE Maturase K.
GN Name=matk;
OS Vanroyenella plumosa.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids I; Malpighiales; Podostemaceae; Vanroyenella.
OX NCBI_TaxID=51609;
[1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048378; BAB3398.1; -.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0009507; P:RNA splicing; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast.
SQ SEQUENCE 508 AA; 61288 MW; 90229C45138AF4B6 CRC64;

Query Match
Best Local Similarity 52.6%; Score 60; DB 2; Length 508;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPOFEK 28
Db 168 WLKDPASLHFLRLEFHEYMNWSPSEK 195

RESULT 14
O9BBG3 PRELIMINARY; PRT; 508 AA.
AC O9BBG3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Maturase K.
GN Name=matk;
OS Oseya coulteriana.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

```

```

OC euroside I; Malpighiales; Podostemaceae; Oseya.
OX NCBI_TaxID=51602;
[1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048375; BAB3395.1; -.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0009507; P:RNA splicing; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast.
SQ SEQUENCE 508 AA; 61400 MW; 45240C7D31B54CDF CRC64;

Query Match
Best Local Similarity 52.6%; Score 60; DB 2; Length 508;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPOFEK 28
Db 168 WLKDPASLHFLRLEFHEYMNWSPSEK 195

RESULT 15
O83DG9 PRELIMINARY; PRT; 650 AA.
AC O83DG9;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Acetoacetyl-CoA synthase.
DE OrderedLocustNames=CBU0766;
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
[1]
RP SEQUENCE FROM N.A.
RA STRAIN=Nine Mile phase I / RSA 493;
RC MEDLINE=22608657; PubMed=12704232; DOI=10.1073/pnas.0931379100;
RX Neishadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettein H., Davidsen T.M., Beaman M.J.,
RA Deboy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
burnetii.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
CC -1-SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC EMBL: AE016962; AAC09302.1; -.
DR TIGR: CB0766; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0016405; F:CoA-ligase activity; IEA.
DR GO: GO:0008299; P:isoprenoid biosynthesis; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR005914; Acac_CoA_synth.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
DR TIGRFAMs: TIGR01217; ac_ac_CoA_synth. 1.
DR PROSITE: PS00455; AMP_BINDING; 1.
DR Complete proteome.
SQ SEQUENCE 650 AA; 72925 MW; FEDEC2ACB51C45A CRC64;

Query Match
Best Local Similarity 52.6%; Score 60; DB 2; Length 650;
Matches 6; Conservative 17; Mismatches 5; Indels 0; Gaps 0;

```

```

Oy      1 WSHPOFEKXXXXXXXXXXXXWSHPOFEK 28
         | : | : : : : : : : : : : | : |
Db      481 WNDPEGKRYQAYFDKYPNTWAHGDAYK 508

```

Search completed: March 2, 2005, 12:44:32
Job time : 106.033 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 126.114 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: seq16

Perfect score: 114
Sequence: 1 whnpqfexkxxxxxxxxxxwhnpqfek 28

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 10%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp19808:*\n2: geneseqp19905:*\n3: geneseqp20008:*\n4: geneseqp20018:*\n5: geneseqp20025:*\n6: geneseqp20038:*\n7: geneseqp20039:*\n8: geneseqp20046:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.5	88.2	35	6	ABP60369 Abp60369 Streptavi
2	100	87.7	36	6	ABP60370 Abp60370 Streptavi
3	98	86.0	24	6	ABP60362 Abp60362 Streptavi
4	69	60.5	24	6	ABP60363 Abp60363 Streptavi
5	68	59.6	763	7	ADH85492 Human ags
6	67.5	59.2	485	8	ADS20249 Furin-pro
7	67.5	59.2	697	8	ADS20227 Human ags
8	66	57.9	179	5	AAE24889 Aae24889 Pneumococ
9	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
10	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
11	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
12	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
13	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
14	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
15	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
16	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
17	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
18	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
19	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
20	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
21	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
22	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
23	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
24	66	57.9	183	5	AAE24889 Aae24889 Pneumococ
25	66	57.9	183	5	AAE24889 Aae24889 Pneumococ

26	65	57.0	330	8	ADN36408 ADN36408 Human pro
27	65	57.0	396	2	AAW93966 Aaw93966 Plasmid p
28	65	57.0	396	4	AAW93966 Aaw93966 Plasmid p
29	65	57.0	396	4	AAW93966 Aaw93966 Plasmid p
30	65	57.0	400	6	ADA27291 Plasmid p
31	65	57.0	400	6	ADA27291 Plasmid p
32	65	57.0	400	6	ADA27291 Plasmid p
33	65	57.0	400	6	ADA27291 Plasmid p
34	65	57.0	400	6	ADA27291 Plasmid p
35	65	57.0	400	6	ADA27291 Plasmid p
36	65	57.0	400	6	ADA27291 Plasmid p
37	65	57.0	400	6	ADA27291 Plasmid p
38	65	57.0	400	6	ADA27291 Plasmid p
39	65	57.0	400	6	ADA27291 Plasmid p
40	65	57.0	400	6	ADA27291 Plasmid p
41	65	57.0	400	6	ADA27291 Plasmid p
42	65	57.0	400	6	ADA27291 Plasmid p
43	65	57.0	400	6	ADA27291 Plasmid p
44	65	57.0	400	6	ADA27291 Plasmid p
45	65	57.0	400	6	ADA27291 Plasmid p

ALIGNMENTS

RESULT 1	ABP60369	standard, peptide, 35 AA.
ID	ABP60369	standard, peptide, 35 AA.
AC	ABP60369	standard, peptide, 35 AA.
XX	ABP60369	standard, peptide, 35 AA.
DT	28-MAR-2003	(first entry)
DE	Streptavidin binding peptide SEQ ID NO 10.	
XX	Streptavidin; protein chip; microtitre plate; detection.	
KW	Synthetic.	
OS	Synthetic.	
FH	Key	Location/Qualifiers
FT	Misc-difference 9..27	/label= unknown
FT	FT	/note= "optionally deleted for 1-15 residues"
XX	DE1013776-A1.	
XX	02-OCT-2002.	
XX	21-MAR-2001; 2001DE-01013776.	
XX	21-MAR-2001; 2001DE-01013776.	
XX	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX	Schmidt T;	
XX	WPI; 2003-031166/03.	
XX	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
XX	Claim 7, Page 16, 18pp; German.	
XX	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it	

CC may not be essential to remove (t); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX
 SQ Sequence 35 AA;

Query Match Best Local Similarity 88.2%; Score 100.5; DB 6; Length 35;
 Matches 28; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 28
 DB 1 WSHPOFEKXXXXXXXXXXXXXXXXXXXXXWSHPOFEK 35

RESULT 2

ABP60370 ID ABP60370 standard; peptide; 36 AA.

AC ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

Key Location/Qualifiers

FT Region 9..28

FT /note= "GGGS repeats 2-5 optionally absent, residues 13-

FT 28" /label= GGGS_repeat

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX Schmidt T;

PI WPI; 2003-031166/03.

DR WPI; 2003-031166/03.

XX New isolated peptide, useful as affinity purification tag for recombinant

PT protein, comprises at least two high-affinity streptavidin-binding

PT modules.

XX Claim 8; Page 16; 18pp; German.

The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 XX

SQ Sequence 36 AA;

Query Match Best Local Similarity 87.7%; Score 100; DB 6; Length 36;
 Matches 16; Conservative 12; Mismatches 0; Indels 8; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 28
 DB 1 WSHPOFEKGGSGGGSGGGSGGGSGGSGWSHPOFEK 36

RESULT 3

ABP60362 ID ABP60362 standard; peptide; 24 AA.

AC ABP60362;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 3.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

Key Location/Qualifiers

FT Misc-difference 9..16

FT /label= unknown

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX Schmidt T;

PI WPI; 2003-031166/03.

DR WPI; 2003-031166/03.

XX The invention relates to an isolated peptide (I) comprising at least two

CC individual modules separated by 0-50 amino acids, with each containing at

CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain

CC streptavidin binding modules, are useful as affinity handles for

CC purification of recombinant fusion proteins (FP), also for detecting FP,

CC e.g. on protein chips or microtitre plates. The modules in (I) bind

CC strongly to streptavidin, with a co-operative effect that provides

CC stronger binding than a single tag, but are displaced by a competitor.

CC (I) does not interfere with the function of attached proteins (II) (so it

CC may not be essential to remove it); facilitates detection and has easily

CC controllable binding properties. (I) is particularly used for purifying

CC FP from dilute solution in batch formats (which use simpler apparatus

CC than column methods and result in lower loss of FP). The present sequence

CC is that of a streptavidin binding peptide disclosed with the invention
 XX

SQ Sequence 24 AA;

Query Match Best Local Similarity 86.0%; Score 98; DB 6; Length 24;
 Matches 24; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXXXXWSHPOFEK 28
 DB 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 24

```

RESULT 4
ABP60363
ID ABP60363 standard; peptide; 24 AA.
XX
AC ABP60363;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 4.
XX
KM Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 9..18
FT /label= unknown
FT Misc-difference 22..24
FT /label= unknown
XX
DE10113776-A1.
XX
PN 02-OCT-2002;
XX
PD 21-MAR-2001; 2001DE-01013776.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
XX
PS Disclosure; Page 4; 18pp; German.
XX
CC The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
XX
SQ Sequence 24 AA;
XX
Query Match 60.5%; Score 69; DB 6; Length 24;
Best Local Similarity 84.0%; Pred. No. 0.15;
Matches 21; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
XX
QY 1 WSHPPQFEKXXXXXXXXXXXXXSHPPQ 25
1 |||||
1 WSHPPQFEKXXXXXXXXXXXXX---HPQ 21
1 |||||
XX
DB 1 WSHPPQFEKXXXXXXXXXXXXX---HPQ 21
XX
RESULT 5
ADBB5492
ID ADBB5492 standard; protein; 763 AA.
XX
AC ADBB5492;
XX

```

```

XX
DT 04-DEC-2003 (first entry)
XX
DE Human aggrecanase-2 (ADAMTS-5) recombinant protein sequence.
XX
KM aggrecanase; aggrecan; articular cartilage; proteoglycan;
KM cartilage compressibility; cartilage elasticity; arthritic disease;
KM osteoarthritis; cartilage degradation; inflammatory joint disease;
KM aggrecanase-1; ADAMTS-4; aggrecanase-2; ADAMTS-5; thrombospondin domain;
KM TSP domain; osteopathic; antiarthritic; cytosolic; antiinflammatory;
KM antirheumatic; ophthalmological; thrombolytic; vasotropic; antimicrobial;
KM respiratory-gen; neuroprotective; antiparkinsonian;
KM immunosuppressive; aggrecanase inhibition; cancer; rheumatoid arthritis;
KM septic arthritis; corneal ulceration; coronary thrombosis;
KM Crohn's disease; emphysema; Alzheimer's disease; Parkinson's disease;
KM multiple sclerosis; aortic aneurysm; enzyme; human.
XX
OS Synthetic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 756..763
FT /label= Streptavidin_tag
XX
PN WO2003066822-A2.
XX
PD 14-AUG-2003.
XX
PF 05-FEB-2003; 2003WO-US003554.
XX
PR 05-FEB-2002; 2002US-0354592P.
XX
PA (AMHP ) WYETH.
XX
PI Georgiadis K, Crawford TK, Tomkinson KV, Morris EA, Racie LJ;
XX
DR WPI: 2003-731495/69.
XX
DR N-PSDB; ADBB5491.
XX
PT New biologically-active aggrecanase protein having a deletion of all, or
PT a portion of a TSP domain, useful for treating osteoarthritis, cancer,
PT Parkinson's disease, coronary thrombosis, Alzheimer's disease and
PT multiple sclerosis.
XX
PS Disclosure; Fig 15; 11pp; English.
XX
CC This invention relates to novel truncated human aggrecanase proteins and
CC nucleotide sequences. Aggrecan is a major extracellular component of
CC articular cartilage. It is a proteoglycan responsible for providing
CC cartilage with its mechanical properties of compressibility and
CC elasticity. The loss of aggrecan has been implicated in the degradation
CC of articular cartilage in arthritic diseases such as osteoarthritis.
CC Aggrecanase is responsible for the cleavage of aggrecan, thereby having a
CC role in cartilage degradation associated with osteoarthritis and
CC inflammatory joint disease. The proteins of the current invention are
CC truncated aggrecanase-1 (ADAMTS-4) and aggrecanase-2 (ADAMTS-5) enzymes
CC which have at least one thrombospondin (TSP) domain deleted. These are
CC biologically active and have greater stability and higher expression than
CC their full-length counterparts. The proteins of the invention may be of
CC use in the development of compounds with osteopathic, antiarthritic,
CC cytostatic, antiinflammatory, antirheumatic, ophthalmological,
CC thrombolytic, vasotropic, antimicrobial, respiratory-gen, neurotropic,
CC neuroprotective, antiparkinsonian or immunosuppressive activities through
CC aggrecanase inhibition. The proteins of the invention may therefore be
CC useful for the manufacture of compositions for the treatment of
CC aggrecanase-associated conditions, such as osteoarthritis, cancer,
CC inflammatory joint disease, rheumatoid arthritis, septic arthritis,
CC corneal ulceration, coronary thrombosis, Crohn's disease, emphysema,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis and aortic
CC aneurysm. The present sequence is the amino acid sequence of a
CC recombinant truncated aggrecanase 2 (ADAMTS-5) enzyme of the invention
CC with a peptide linker and a streptavidin tag.
XX

```

SQ Sequence 763 AA;
 Query Match 59.6%; Score 68; DB 7; Length 763;
 Best Local Similarity 42.9%; Pred. No. 37;
 Matches 9; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
 QY 8 KXXXXXXXXXXXXXSHPOPEK 28
 |:::|||||
 Db 743 KGYTDVVRIPGSAMSHPOPEK 763
 RESULT 6
 ADS20249
 ID ADS20249 standard; protein; 485 AA.
 AC ADS20249;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Furin-processed human aggrecanase ADAMTS4 truncated protein w Strept tag.
 KW ADAMTS4; a disintegrin-like and metalloprotease;
 KW chondrospondin type 1 motif 4; reprolysin; zinc metalloprotease;
 KW aggrecanase; osteopontin; antiinflammatory; antiarthritic; antineumatic;
 KW cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
 KW human; chromosome 1q21-q23; enzyme; truncation; mature; furin cleavage;
 KW Strept tag.
 XX
 OS Homo sapiens.
 OS Unidentified.
 FH Key Location/Qualifiers
 FT Misc-difference 475..485
 FT /note="Residues corresponding to positions 687-837 in
 FT the wild-type replaced by Strept tag"
 FT
 XX
 PN WO2004011637-A2.
 XX
 PD 05-FEB-2004.
 XX
 PF 29-JUL-2003; 2003WO-US023484.
 XX
 PR 29-JUL-2002; 2002US-0398721P.
 XX
 PA (AMHP) WYETH.
 PA (CORC) CORCORAN C J.
 PA (FLAN) FLANNERY C R.
 PA (ZENG) ZENG W.
 PA (RACI) RACIE L A.
 PA (MCDO) MCDONAGH T.
 PA (FREE) FREEMAN B A.
 PA (GEOR) GEORGIADIS K E.
 PA (LAVA) LAVALLIE E R.
 PA
 PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
 PI Freeman BA, Georgiadis KE, Lavallie ER;
 PI WPI; 2004-143860/14.
 DR
 XX
 PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
 PT stability useful for identifying inhibitors of the enzyme activity for
 PT treating aggrecanase-associated conditions, including osteoarthritis.
 PT
 XX
 PS Claim 9; SEQ ID NO 47; 117pp; English.
 XX
 CC The invention relates to a novel isolated, modified ADAMTS4 (a
 CC disintegrin-like and metalloprotease (reprolysin type) with
 CC chondrospondin type 1 motif 4) protein with improved stability compared
 CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
 CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
 CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc
 CC metalloproteases and include aggrecanases amongst their members. The

CC protein of the invention demonstrates osteopontin, antiinflammatory,
 CC antirheumatic, antineumatic and cytostatic activities and may be useful
 CC for treating aggrecanase-associated conditions, including osteoarthritis,
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
 CC arthritis, periodontal disease and Crohn's disease. The current sequence
 CC is that of the furin-processed human aggrecanase ADAMTS4 truncated
 CC protein with Strept tag of the invention.
 XX
 SQ Sequence 485 AA;
 Query Match 59.2%; Score 67.5; DB 8; Length 485;
 Best Local Similarity 36.7%; Pred. No. 22;
 Matches 11; Conservative 14; Mismatches 2; Indels 3; Gaps 1;
 QY 2 SHPOPEK--XXXXXXXXXXXXXSHPOPEK 28
 |:::|||||
 Db 456 SKKFKCMVCGGDGGSGSAMSHPOPEK 485
 RESULT 7
 ADS20227
 ID ADS20227 standard; protein; 697 AA.
 AC ADS20227;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human aggrecanase ADAMTS4 truncated protein with Strept tag - SEQ ID 24.
 KW ADAMTS4; a disintegrin-like and metalloprotease;
 KW chondrospondin type 1 motif 4; reprolysin; zinc metalloprotease;
 KW aggrecanase; osteopontin; antiinflammatory; antiarthritic; antineumatic;
 KW cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
 KW human; enzyme; chromosome 1q21-q23; truncation; Strept tag.
 XX
 OS Homo sapiens.
 OS Unidentified.
 FH Key Location/Qualifiers
 FT Misc-difference 687..697
 FT /note="Wild-type residues 687-837 replaced by Strept tag"
 FT
 XX
 PN WO2004011637-A2.
 XX
 PD 05-FEB-2004.
 XX
 PF 29-JUL-2003; 2003WO-US023484.
 XX
 PR 29-JUL-2002; 2002US-0398721P.
 XX
 PA (AMHP) WYETH.
 PA (CORC) CORCORAN C J.
 PA (FLAN) FLANNERY C R.
 PA (ZENG) ZENG W.
 PA (RACI) RACIE L A.
 PA (MCDO) MCDONAGH T.
 PA (FREE) FREEMAN B A.
 PA (GEOR) GEORGIADIS K E.
 PA (LAVA) LAVALLIE E R.
 PA
 PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
 PI Freeman BA, Georgiadis KE, Lavallie ER;
 PI WPI; 2004-143860/14.
 DR
 XX
 PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
 PT stability useful for identifying inhibitors of the enzyme activity for
 PT treating aggrecanase-associated conditions, including osteoarthritis.
 PT
 XX
 PS Claim 9; SEQ ID NO 24; 117pp; English.
 XX
 CC The invention relates to a novel isolated, modified ADAMTS4 (a


```

PN  DE10025620-A1.
XX
PD  06-DEC-2001.
XX
PF  24-MAY-2000; 2000DE-01025620.
XX
PR  24-MAY-2000; 2000DE-01025620.
XX
PA  (GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
XX
PI  Pongs O, Legros C, Eaucilaire M, Bougis PE;
XX
DR  WPI; 2002-042513/06.
XX
DR  N-PSDB; AAI69255.
XX
PT  New nucleic acid encoding chimeric potassium channel protein, useful,
XX  when immobilized on carrier, in screening agents for toxicity and
XX  pharmaceutical activity.
XX
PS  Example 1, Page 15; 34pp; German.
XX
CC  This invention describes a novel chimeric nucleic acid (I'), encoding a
CC  potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC  and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC  useful, after fixing to a carrier, for identifying compounds that bind to
CC  the chimeras or modify their activity, i.e. compounds that are toxic or
CC  potentially useful therapeutically by acting as potassium channel
CC  antagonist or agonist, in treatment of circulatory or nervous system
CC  disorders, urinary incontinence, tumours, or inflammation, in human or
CC  veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC  as an array, on a carrier, making possible screening by protein chip
CC  technology. The channels can be isolated from recombinant bacteria in
CC  stable, active and functional form, and membranes are stable for many
CC  weeks or months, during which time they can be used many times. This
CC  sequence represents Chivi subunit used in the construction of the
CC  chimeric potassium channels described in the invention
XX
SQ  Sequence 183 AA;

Query Match          57.9%; Score 66; DB 5; Length 183;
Best Local Similarity 29.2%; Pred. No. 8;
Matches 7; Conservative 15; Mismatches 2; Indels 0; Gaps 0.

QY  3 HQFEKXXXXXXXXXXXXWSHPQF 26
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  158 HERFRLERLDDNRRSAWRHPQF 181

RESULT 10
AAG80261
ID  AAG80261 standard; protein; 183 AA.
XX
AC  AAG80261;
XX
DE  18-FEB-2002. (first entry)
XX
DE  Chimeric potassium channel KcsA-Kv1/3 ChiviI.
XX
KW  Potassium channel; chimeric; KcsA, Kv3.1; agonist; antagonist; tumour;
KW  circulatory disorder; nervous system disorder; urinary incontinence;
XX  inflammation; protein chip; screening; ChiviI.
XX
XX  Homo sapiens.
XX  OS  Streptomyces lividans.
XX  OS  Synthetic.
XX
XX  DE10025620-A1.
XX
XX  06-DEC-2001.
XX
XX  24-MAY-2000; 2000DE-01025620.
XX
XX  24-MAY-2000; 2000DE-01025620.
XX
XX  24-MAY-2000; 2000DE-01025620.
XX

```

```

XX XX (GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
PA Pongs O, Legros C, Eauclaire M, Bougis PE;
XX WPI; 2002-042513/06.
DR N-PSDB; AA169256.
XX New nucleic acid encoding chimeric potassium channel protein, useful,
PT when immobilized on carrier, in screening agents for toxicity and
PT pharmaceutical activity.
XX Example 1; Page 16; 34pp; German.
XX This invention describes a novel chimeric nucleic acid (I), encoding a
CC potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC useful, after fixing to a carrier, for identifying compounds that bind to
CC the chimeras or modify their activity, i.e. compounds that are toxic or
CC potentially useful therapeutically, by acting as potassium channel
CC antagonist or agonist, in treatment of circulatory or nervous system
CC disorders, urinary incontinence, tumours, or inflammation, in human or
CC veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC as an array, on a carrier, making possible screening by protein chip
CC technology. The channels can be isolated from recombinant bacteria in
CC stable, active and functional form, and membranes are stable for many
CC weeks or months, during which time they can be used many times. This
CC sequence represents the ChlVI subunit used in the construction of the
CC chimeric potassium channels described in the invention
SQ Sequence 183 AA;
Query Match 57.9%; Score 66; DB 5; Length 183;
Best Local Similarity 29.2%; Pred. No. 8;
Matches 7; Conservative 15; Mismatches 2; Indels 0; Gaps 0;
QY 3 HQPEKXXXXXXXXXXXXXWHPDF 26
| : ::::::::::::::: | ||||
Db 158 HERFDRLRMLDNNRRSAMRHPOF 181
RESULT 11
AAG80255
ID AAG80255 standard; protein; 183 AA.
XX
AC AAG80255;
DT 18-FEB-2002 (first entry)
DE Chimeric potassium channel KcsA-Kv1.3 ChII.
XX
KW Potassium channel; chimeric; KcsA; Kv1.3; agonist; antagonist; tumour;
KW circulatory disorder; nervous system disorder; urinary incontinence;
KW inflammation; protein chip; screening; ChII.
XX Homo sapiens.
OS Streptomyces lividans.
XX Synthetic.
XX DE10025620-A1.
XX
XX 06-DEC-2001.
XX
XX 24-MAY-2000; 2000DE-01025620.
XX
XX 24-MAY-2000; 2000DE-01025620.
XX
XX (GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
PA Pongs O, Legros C, Eauclaire M, Bougis PE;
XX WPI; 2002-042513/06.
DR N-PSDB; AA169250.

```

CC This invention describes a novel chimeric nucleic acid (I), encoding a
CC potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC useful, after fixing to a carrier, for identifying compounds that bind to
CC the chimera or modify their activity, i.e. compounds that are toxic or
CC potentially useful therapeutically, by acting as potassium channel
CC antagonist or agonist, in treatment of circulatory or nervous system
CC disorders, urinary incontinence, tumours, or inflammation, in human or
CC veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC as an array, on a carrier, making possible screening by protein chip
CC technology. The channels can be isolated from recombinant bacteria in
CC stable, active and functional form, and membranes are stable for many
CC weeks or months, during which time they can be used many times. This
CC sequence represents the ChlV subunit used in the construction of the
CC chimeric potassium channels described in the invention

CC CC
CC Sequence 183 AA;
SQ

Query Match 57.9%; Score 66; DB 5; Length 183;
Best Local Similarity 29.2%; Pred. No. 8;
Matches 7; Conservative 15; Mismatches 2; Indels 0; Gaps 0;

Oy 3 HPOFEKXXXXXXXXXXWSHPQF 26
 |:|:|||||||
Db 158 HERFDRLRMDDNRRSAMRHHQPF 181

RESULT 13
AAG80263
ID AAG80263 standard; protein; 183 AA.
XX AC
XX AAC80263;
DT 18-FEB-2002 (first entry)
DE Chimeric potassium channel KcsA-Kv1/3 Ch1X.
XX DE
XX Potassium channel; chimeric; KcsA; Kv3.1; agonist; antagonist; tumour;
KW circulatory disorder; nervous system disorder; urinary incontinence;
KM inflammation; protein chip; screening; Ch1X.
OS Homo sapiens.
OS Streptomyces lividans.
OS Synthetic.
XX OS
FH Key Location/Qualifiers
FT Misc-difference 181 /note= "Encoded by AA"
XX FT
XX DE10025620-A1.
PN PN
PD PD
PP 06-DEC-2001.
PX PX
PY 24-MAY-2000; 2000DB-01025620.
PR PR
PS 24-MAY-2000; 2000DB-01025620.
PA (GENE-) FORSCHUNGSGESSELLSCHAFT GENION MBH.
PI Pongs O, Legros C, Baucilaire M, Bougis PE,
DR WPI; 2002-042513/06.
DR N-PADB; AAI69258.
XX XX
XX New nucleic acid encoding chimeric potassium channel protein, useful,
PT when immobilized on carrier, in screening agents for toxicity and
XX pharmaceutical activity.
XX PS
PS Example 2; Page 17; 34pp; German.

CC This invention describes a novel chimeric nucleic acid (I), encoding a
CC potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC useful, after fixing to a carrier, for identifying compounds that bind to
CC the chimera or modify their activity, i.e. compounds that are toxic or
CC potentially useful therapeutically, by acting as potassium channel
CC antagonist or agonist, in treatment of circulatory or nervous system
CC disorders, urinary incontinence, tumours, or inflammation, in human or
CC veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC as an array, on a carrier, making possible screening by protein chip
CC technology. The channels can be isolated from recombinant bacteria in
CC stable, active and functional form, and membranes are stable for many
CC weeks or months, during which time they can be used many times. This
CC sequence represents the ChIV subunit used in the construction of the
CC chimeric potassium channels described in the invention

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 87.0732 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ16
Perfect score: 114
Sequence: 1 wehpqfexxxxxxxxxxxwehpqfex 28

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubppa/US09C_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubppa/US10D_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	87.7	36	14	US-10-026-578B-10
2	98	86.0	24	14	US-10-026-578B-3
3	87	76.3	36	14	US-10-026-578B-11
4	77	67.5	24	9	US-09-809-517A-31
5	77	67.5	25	9	US-09-809-517A-34
6	77	67.5	245	17	US-10-887-228A-1
7	77	67.5	246	17	US-10-887-228A-9
8	77	67.5	252	17	US-10-887-228A-5
9	69	60.5	24	14	US-10-026-578B-4
10	68	59.6	763	15	US-10-358-283-15
11	67.5	59.2	485	16	US-10-628-432-47
12	67.5	59.2	697	16	US-10-628-432-24
13	65	57.0	646	16	US-10-628-432-49

14	65	57.0	845	16	US-10-628-432-40	Sequence 40, Appl
15	65	57.0	858	16	US-10-628-432-27	Sequence 27, Appl
16	64.5	56.6	158	15	US-10-424-599-184769	Sequence 184769,
17	64.5	56.6	212	15	US-10-424-599-184075	Sequence 184075,
18	64	56.1	21	9	US-09-809-517A-30	Sequence 30, Appl
19	63.5	55.7	22	9	US-09-809-517A-33	Sequence 33, Appl
20	63.5	55.7	633	16	US-10-628-432-53	Sequence 53, Appl
21	63	55.3	117	10	US-09-977-137A-4	Sequence 4, Appl
22	63	55.3	117	10	US-09-977-137A-5	Sequence 5, Appl
23	63	55.3	117	10	US-09-977-137A-7	Sequence 7, Appl
24	63	55.3	117	10	US-09-977-137A-8	Sequence 8, Appl
25	63	55.3	117	10	US-09-977-137A-9	Sequence 9, Appl
26	63	55.3	117	10	US-09-977-137A-10	Sequence 10, Appl
27	63	55.3	117	10	US-09-977-137A-11	Sequence 11, Appl
28	63	55.3	117	10	US-09-977-137A-12	Sequence 12, Appl
29	63	55.3	118	10	US-09-977-137A-6	Sequence 6, Appl
30	63	55.3	661	15	US-10-354-983-8	Sequence 8, Appl
31	61	53.5	621	15	US-10-282-122A-50515	Sequence 50515, A
32	60.5	53.1	134	16	US-10-437-963-104065	Sequence 104065,
33	60.5	53.1	252	16	US-10-437-963-200150	Sequence 200150,
34	60	52.6	102	16	US-10-767-701-47647	Sequence 47647, A
35	60	52.6	291	15	US-10-425-114-50330	Sequence 50330, A
36	60	52.6	305	16	US-10-767-701-42846	Sequence 42846, A
37	60	52.6	309	15	US-10-425-114-61196	Sequence 61196, A
38	60	52.6	309	15	US-10-425-114-65584	Sequence 65584, A
39	60	52.6	309	15	US-10-425-114-71583	Sequence 71583, A
40	60	52.6	311	15	US-10-425-114-62543	Sequence 62543, A
41	60	52.6	311	15	US-10-425-114-70843	Sequence 70843, A
42	59	51.8	109	16	US-10-378-029-65	Sequence 65, Appl
43	59	51.8	168	15	US-10-437-963-172987	Sequence 172987,
44	59	51.8	759	16	US-10-437-963-118124	Sequence 118124,
45	59	51.8	1527	16	US-10-437-963-201843	Sequence 201843,

ALIGNMENTS

RESULT 1
US-10-026-578B-10
Sequence 10, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(28)
OTHER INFORMATION: X represents a single amino acid at each of the positions indicat
OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are
OTHER INFORMATION: missing, the total numbers of x will be no less than 5
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10
Query Match 87.7%; Score 100; DB 14; Length 36;
Best Local Similarity 77.8%; Pred. No. 6.6e-05;

```

Query Match      86.0%; Score 98; DB 14; Length 24;
Best Local Similarity 85.7%; Pzed. No. 6.9e-05;
Matches 24; Conservative 0; Mismatches 0; Indels 4; Gaps 1
QY 1 WSHPOFEKXXXXXXXXXXXXXSHPOFEK 28
||||||| |||||||||||||||

```

```

RESULT 4
US-09-809-517A-31
/ Sequence 31, Application US/09809517A
/ Patent No. US20020034733A1
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/proteins
/ TITLE OF INVENTION: Particles via disulfide bonds
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ CURRENT FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ PRIOR FILING DATE: 2000-02-18
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 31
/ LENGTH: 24
/ TYPE: PRT
/ ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

```

Query Match	67.5%	Score 77;	DB 9;	Length 24;
Best Local Similarity	41.7%;	Pred. No. 0.033;		

```
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 1 EFQKLISEBDLNGAPWSPQPEK 24

RESULT 5
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match 67.5%; Score 77; DB 9; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.035;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 2 EFQKLISEBDLNGAPWSPQPEK 25

RESULT 6
US-10-887-228A-1
; Sequence 1, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S305690S
; CURRENT APPLICATION NUMBER: US/10/887,228A
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 245
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-1

Query Match 67.5%; Score 77; DB 17; Length 245;
Best Local Similarity 41.7%; Pred. No. 0.74;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 229 EFQKLISEBDLNGAPWSPQPEK 246

RESULT 7
US-10-887-228A-9
; Sequence 9, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S305690S
; CURRENT APPLICATION NUMBER: US/10/887,228A
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 246
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match 67.5%; Score 77; DB 17; Length 246;
Best Local Similarity 41.7%; Pred. No. 0.74;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 223 EFQKLISEBDLNGAPWSPQPEK 246

RESULT 8
US-10-887-228A-5
; Sequence 5, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S305690S
; CURRENT APPLICATION NUMBER: US/10/887,228A
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 252
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-5

Query Match 67.5%; Score 77; DB 17; Length 252;
Best Local Similarity 41.7%; Pred. No. 0.77;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 229 EFQKLISEBDLNGAPWSPQPEK 252

RESULT 9
US-10-026-578B-4
; Sequence 4, Application US/10026578B
```

```

Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
PRIORITY FILING DATE: 2002-11-11
PRIORITY APPLICATION NUMBER: DE 101 13 776.1
PRIORITY FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: PCT/EP01/11846
PRIORITY FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC_FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (10)..(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (12)..(12)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (13)..(13)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (14)..(14)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (15)..(15)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (16)..(16)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (17)..(17)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (18)..(18)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (19)..(19)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (20)..(20)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (21)..(21)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (22)..(22)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (23)..(23)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (24)..(24)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:

```

```

NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-4

```

```

Query Match 60.5%; Score 69; DB 14; Length 24;
Best Local Similarity 84.0%; Pred. No. 0.34;
Matches 21; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

```

```

QY 1 WSHPOFEKXXXXXXXXXXWSHPQ 25
DB 1 WSHPOFEKXXXXXXXXXX----HPQ 21

```

```

RESULT 10
US-10-358-283-15
Sequence 15, Application US/10358283
Publication No. US2004005419A1
GENERAL INFORMATION:
APPLICANT: WYETH
TITLE OF INVENTION: TRUNCATED AGGREGINASE MOLECULES
FILE REFERENCE: 08702-0112-00000
CURRENT APPLICATION NUMBER: US/10/358,283
PRIORITY FILING DATE: 2003-02-17
PRIORITY APPLICATION NUMBER: 60/354,592
PRIORITY FILING DATE: 2002-02-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 763
TYPE: PRT
ORGANISM: Homo sapiens
US-10-358-283-15

```

```

Query Match 59.6%; Score 68; DB 15; Length 763;
Best Local Similarity 42.9%; Pred. No. 48;
Matches 9; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 8 KXXXXXXXXXXWSHPQFEK 28
DB 743 KGYTDVVRIPGSAMSHPOFEK 763

```

```

RESULT 11
US-10-628-432-47
Sequence 47, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
PRIORITY FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47
LENGTH: 485
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: furin-processed construct C
US-10-628-432-47

```

```

Query Match 59.2%; Score 67.5; DB 16; Length 485;
Best Local Similarity 36.7%; Pred. No. 30;
Matches 11; Conservative 14; Mismatches 2; Indels 3; Gaps 1;

```

```

QY 2 SHPOFEK---XXXXXXXXXXWSHPQFEK 28
DB 456 SKKFDKCMVCGDGGSGSGSAMSHPOFEK 485

```

```

RESULT 12
US-10-628-432-24

```

```

Sequence 24 Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 697
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Truncated ADAMTS4 molecule
US-10-628-432-24

Query Match          59.2%; Score 67.5; DB 16; Length 697;
Best Local Similarity 36.7%; Pred. No. 49;
Matches 11; Conservative 14; Mismatches 2; Indels 3; Gaps 1

Oy      2 SHPOFEK---XXXXXXXXXXXXXWSHPOFEK 28
        |||::|| ::::::::::::::|||
Db       668 SKKFPDKMVCGGDSCGSGSAMSHPOFEK 697

RESULT 13
US-10-628-432-49,
Sequence 49, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49
LENGTH: 646
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: furin-processed construct E
US-10-628-432-49

Query Match          57.0%; Score 65; DB 16; Length 646;
Best Local Similarity 31.2%; Pred. No. 92;
Matches 10; Conservative 14; Mismatches 4; Indels 4; Gaps 1

Oy      1 WSHPOFE-----KXXXXXXXXXXXXXWSHPOFEK 28
        |||::|| ::::::::::::::|||
Db       615 WLHRRQAILELIRRPWAGRKSAWSHPOFEK 646

RESULT 14
US-10-628-432-40
Sequence 40, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 40
LENGTH: 845
TYPE: PRT
ORGANISM: Artificial
FEATURE:

```

```

; OTHER INFORMATION: ADAMTS4 ASM with insertion
US-10-628-432-40

Query Match          57.0%; Score 65; DB 16; Length 845;
Best Local Similarity 40.9%; Pred. No. 1.3e+02;
Matches 9; Conservative 12; Mismatches 1; Indels 0; Gaps 0;

Oy      1 WSHPOFEKXXXXXXXXXXWS 22
        |||||:||||:||||:|
Db      520 WSHPOFEKAGCGPWPWGDCDS 541

RESULT 15
US-10-628-432-27
; Sequence 27, Application US/10628432
; Publication NO. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

Query Match          57.0%; Score 65; DB 16; Length 858;
Best Local Similarity 31.2%; Pred. No. 1.3e+02;
Matches 10; Conservative 14; Mismatches 4; Indels 4; Gaps 1;

Oy      1 WSHPOFE---KXXXXXXXXXXWSHPQFEK 28
        |||:|:||||:||||:|||||
Db      827 WLRRAQLLEILRRRPPWAGRGSAWSHPQFEK 858

Search completed: March 2, 2005, 14:19:04
Job time : 88.0732 secs

```

This Page Blank (uspto)


```
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 34
/ LENGTH: 25
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```
Query Match          67.5%; Score 77; DB 4; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.0014;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 OEQXXXXXXXXXXWSHPQFEK 28
Db 2 EFDYKDDDDKGAWSHPQFEK 25
```

```
RESULT 3
US-09-809-517A-30
/ Sequence 30, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
/ FILE REFERENCE: WOPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 30
/ LENGTH: 21
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
```

```
Query Match          56.1%; Score 64; DB 4; Length 21;
Best Local Similarity 38.1%; Pred. No. 0.075;
Matches 8; Conservative 13; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 KXXXXXXXXXXWSHPQFEK 28
Db 1 EFDYKDDDDKGAWSHPQFEK 21
```

```
RESULT 4
US-09-809-517A-33
/ Sequence 33, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 33
/ LENGTH: 22
/ TYPE: PRT
/ ORGANISM: artificial sequence
```

```
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33
```

```
Query Match          56.1%; Score 64; DB 4; Length 22;
Best Local Similarity 38.1%; Pred. No. 0.081;
Matches 8; Conservative 13; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 KXXXXXXXXXXWSHPQFEK 28
Db 2 EFDYKDDDDKGAWSHPQFEK 22
```

```
RESULT 5
US-09-977-137A-4
/ Sequence 4, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4
```

```
Query Match          55.3%; Score 63; DB 4; Length 117;
Best Local Similarity 40.0%; Pred. No. 1.5;
Matches 8; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 9 XXXXXXXXXXXXXWSHPQFEK 28
Db 98 HARKGNVSCPSAWSHPQFEK 117
```

```
RESULT 6
US-09-977-137A-5
/ Sequence 5, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5
```

```
Query Match          55.3%; Score 63; DB 4; Length 117;
Best Local Similarity 40.0%; Pred. No. 1.5;
Matches 8; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      9 XXXXXXXXXXXXXWSHPQFEK 28
          :::::::::::|||||
Db     98 HARKGNVSCPSAWSHPQFEK 117
```

RESULT 7
116-09-97

```

Sequence 7 Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Caguiat, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
TITLE OF INVENTION: Methods
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

```

Query Match	55.3%	Score 63	DB 4	Length 117
Best Local Similarity	40.0%	Pred. No. 1.5		
Matches	8	Conservative	12	Mismatches 0; Indels 0; Gaps 0

```
Qy      9 XXXXXXXXXXXXXWSHPQFEK 28
          ::::::::::|
Db      98 HARKGNVSCPSAWSHPQFEK 117
```

RESULT 8
US-09-977-137A-8

```

? Sequence 8, Application US/09977137A
? Patent No. 6750042
? GENERAL INFORMATION:
? APPLICANT: Summers, Anne O.
? APPLICANT: Caguiat, Jonathan
? TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
? TITLE OF INVENTION: Methods
? FILE REFERENCE: 79-00
? CURRENT APPLICATION NUMBER: US/09/977,137A
? CURRENT FILING DATE: 2001-10-12
? PRIOR APPLICATION NUMBER: US 60/240,465
? PRIOR FILING DATE: 2000-10-12
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 8
? LENGTH: 117
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: chelon
? US-09-977-137A-8

```

Query Match	55.3%	Score 63	DB 4	Length 117
Similarity	40.0%	Pred. NO. 1.5		
Best Local				
Matches	8	Conservative	12	Mismatches 0
				Indels 0
				Gaps 0

```

QY      9 XXXXXXXXXXXXWSHPQFEK 28
          ::::::::::|
          98 HARKGNVSCPSAWSHPOFEK 117
          Db

```

RESULT 9

US-09-977-137A-9
; Sequence 9, Application US/09977137A

Query Match	55.3%	Score 63;	DB 4;	Length 117;
Best Local Similarity	40.0%;	Pred. No. 1.5;		
Matches	8;	Conservative	12;	Mismatches
			0;	Indels
				0;
				Gaps
				0;

```
QY      9 XXXXXXXXXXXXWSHPQFEK 28
          ::::::::::::::::::::
Db     98 HARKGNVSCPSAWSHPQFEK 117
```

RESULT 10
US-09-977-137A-10

```

Sequence 10, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Cogniat, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
TITLE OF INVENTION: Methods
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

```

Query Match	55.3%	Score 63;	DB 4;	Length 117;
Best Local Similarity	40.0%	Pred. NO. 1.5;		
Matches	8;	Conservative	12;	Mismatches
			0;	Indels
				0;
				Gaps
				0;

```
QY      9 XXXXXXXXXXXXXWSHPQFEK 28
          ::::::::::::::::::::
Db      98 HARKGNVSCPSAMWSHPQFEK 117
```

RESULT 11
US-09-977-137A-11

```

: Sequence 11, Application US/08977137A
: Patent No. 6750042
:
: GENERAL INFORMATION:
:
: APPLICANT: Summers, Anne O.
: APPLICANT: Cagniat, Jonathan
: TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

```


Db 14 YKHPQETKSNRRDRAGRLISEWARP1F 39

Search completed: March 2, 2005, 12:25:39
job time : 31.8699 secs

This Page Blank (uspto)